Draft 4/20/04

Debiasing Through Law

Christine Jolls[*]
Cass R. Sunstein[**]

Abstract

A substantial literature in psychology and behavioral economics emphasizes ways in which individuals are boundedly rational. In the face of such bounded rationality, the legal system might attempt either to "debias law," by insulating legal outcomes from the effects of bounded rationality, or instead to "debias through law," by steering legal actors in more rational directions. It is conventional in existing legal analyses to focus almost exclusively on insulating outcomes from the effects of bounded rationality. In fact, however, a large number of actual and imaginable legal strategies can be seen as efforts to engage in debiasing through *law – to help people reduce or even eliminate boundedly rational behavior. In important contexts, these efforts promise to avoid the costs and inefficiencies associated with regulatory approaches that take bounded rationality as a given and respond by attempting to insulate outcomes from its effects. This Article provides examples of debiasing through law from many areas, including consumer safety law, employment discrimination law, wrongful discharge law, property law, and corporate law. In some of these areas, the law shows an implicit behavioral rationality; in others, there are opportunities for promising reforms that give greater emphasis to debiasing through law. Discussion is devoted to the risks of overshooting and manipulation that are sometimes raised when government engages in debiasing through law.*

---

[*] Professor of Law, Harvard Law School.
[**] Karl N. Llewellyn Distinguished Service Professor of Jurisprudence, University of Chicago. For helpful comments, we are very grateful to Bruce Ackerman, Ian Ayres, Lucian Bebchuk, Cary Coglianese, Bert Huang, Louis Kaplow, Eric Posner, Jeffrey Rachlinski, Reva Siegel, Peter Siegelman, Adrian Vermeule, and participants in workshops at Fordham Law School, Harvard Law School, the John F. Kennedy School of Government at Harvard University, and Yale Law School.

Much legal analysis now focuses on ways in which human behavior deviates systematically from what would be predicted by the traditional law and economics assumption of unbounded rationality.[1] To the extent that legal rules are designed partly on the basis of their anticipated effects on behavior, bounded rationality is obviously relevant to the formulation of legal policy. But an important and little addressed question is precisely *how* it is relevant to the formulation of legal policy. The most obvious and most familiar possibility is that, given a demonstration of the existence and importance of a particular aspect of bounded rationality, law should be structured to presume the existence of that particular shortcoming in human behavior.

Examples abound. Consider, for instance, the possibility that optimism bias – the pronounced tendency of individuals to underestimate the likelihood of negative events – leads consumers to assume that potentially risky products are substantially safer than they in fact are.[2] If so, the law might respond by adopting heightened standards of manufacturer liability for consumer products.[3] Or assume that hindsight bias – the tendency of individuals to attach an excessively high probability to an event simply because it ended up occurring – adversely affects judgments reached by ex post decision makers on matters of corporate law.[4] If so, then the law could respond, as indeed it has with the "business judgment rule," by adopting an approach that largely vitiates the liability of corporate law actors, who would otherwise be vulnerable to hindsight-biased judgments on the part of adjudicators.[5] More generally, rules and institutions might be designed so that legal outcomes do not fall prey to problems of bounded rationality. Boundedly rational behavior thus might be taken to justify a strategy of *insulation*, attempting to protect legal outcomes from falling victim to this behavior. To date, virtually all of the treatments of bounded rationality in law

---

[1] See, e.g., Symposium, Empirical Legal Realism: A New Social Scientific Assessment of Law and Human Behavior, 97 Nw. U.L. Rev. 1075 (2003); Christine Jolls, Cass R. Sunstein & Richard Thaler, A Behavioral Approach to Law and Economics, 50 Stan. L. Rev. 1471 (1998).

[2] See Jon D. Hanson & Douglas A. Kysar, Taking Behavioralism Seriously: The Problem of Market Manipulation, 74 N.Y.U. L. Rev. 632, 729-30 (1999).

[3] See Jon D. Hanson & Douglas A. Kysar, Taking Behavioralism Seriously: Some Evidence of Market Manipulation, 112 Harv. L. Rev. 1420, 1560 (1999).

[4] See Jeffrey J. Rachlinski, A Positive Psychological Theory of Judging in Hindsight, 65 U. Chi. L. Rev. 571, 620-21 (1998).

[5] See id.

have been of this character. Strategies for insulation can be characterized as a method for "debiasing law."

A quite different possibility – one that has not received much attention in law or elsewhere – is that legal policy may respond best to problems of bounded rationality not by structuring rules and institutions to protect legal outcomes from the effects of such behavior (which itself is taken as a given), but instead by operating directly on the boundedly rational behavior and attempting to help people either to reduce or to eliminate it. We describe legal policy in this category as "debiasing through law." The promise of strategies for debiasing through law is that as compared with the usual approach of attempting to insulate legal outcomes from the effects of bounded rationality, these strategies will often be a less intrusive, more direct, and more democratic response to the problem of bounded rationality.

In fact there exists a substantial, empirically-oriented social science literature on the debiasing of individuals after a demonstration of the existence of a given form of bounded rationality.[6] But empirical findings on these forms of debiasing have made only limited appearances in the legal literature,[7] and equally important, social scientists interested in such forms of debiasing have not investigated the possibility of achieving them through law. In many important settings, empirical evidence suggests the substantial potential of these sorts of debiasing strategies, and from a legal policy perspective it is obviously important to ask about the role that law can play in facilitating such debiasing. That is our major focus in this Article. Instead of trying to "debias law" through insulating legal outcomes from boundedly rational behavior – which is itself viewed as unavoidable – we are interested in debiasing, through law, individuals who exhibit bounded rationality. We

---

[6] Leading examples include Baruch Fischhoff, Debiasing, in Judgment under Uncertainty 422, 424 (Daniel Kahneman et al. eds., 1982); Lawrence Sanna, Norbert Schwarz & Shavaun L. Stocker, When Debiasing Backfires: Accessible Content and Accessibility Experiences in Debiasing Hindsight, 28 J. Experimental Psychol.: Learning, Memory, & Cognition 497 (2002); and Neil D. Weinstein & William M. Klein, Resistance of Personal Risk Perceptions to Debiasing Interventions, in Heuristics and Biases: The Psychology of Intuitive Judgment 313 (Thomas Gilovich et al. eds., 2002).

[7] As noted in the text just below, where debiasing has been examined in the legal literature, the focus has been on achieving debiasing through procedural rules governing adjudication by judges or juries. We discuss several examples below.

show that some areas of law demonstrate an implicit awareness of opportunities for debiasing through law; in that sense, part of our analysis is descriptive. But we have prescriptive goals as well, attempting to demonstrate that debiasing through law holds out important promise in a number of domains.

When debiasing through law has been discussed in the legal literature, the treatment has focused on existing or proposed steps taken in *procedural* rules governing adjudication by judges or juries. A well-known example of this approach is the work by Linda Babcock, George Loewenstein, and Samuel Issacharoff on self-serving bias.[8] Many litigants appear to exhibit this bias, evaluating likely outcomes, as well as questions of fairness, in ways that are systematically self-serving.[9] Babcock, Loewenstein, and Issacharoff find, however, that in an experimental setting the bias may be eradicated by requiring litigants to consider the weaknesses in their case or reasons that the judge might rule against them.[10] (It is possible that "real" self-serving bias – outside of the lab – is more resilient.[11]) In a similar vein, anchoring – in which judgments are influenced by arbitrary cues presented to decision makers, such as the dollar amount requested in a legal complaint – has been shown to produce important effects on jury awards[12]; in response, procedural reforms might be adopted to ensure that juries consider other facts or features of the case in addition to the anchor.[13] This Article, by contrast, emphasizes a different and broader form of debiasing through law – a category we call "debiasing through substantive law."

---

[8] Linda Babcock, George Loewenstein, & Samuel Issacharoff, Creating Convergence: Debiasing Biased Litigants, 22 L. & Soc. Inquiry 913 (1997)

[9] Id. at 917-18; Seth Seabury & Eric Talley, Private Information, Self-serving Biases, and Optimal Settlement Mechanisms: Theory and Evidence (working paper).

[10] Babcock, Loewenstein, & Issacharoff, supra note 8, at 918-19.

[11] See Ward Farnsworth, The Legal Regulation of Self-Serving Bias, 37 U.C. Davis L. Rev. 567, 582-85 (2003).

[12] Gretchen Chapman & Eric Johnson, Incorporating the Irrelevant: Anchors in Judgments of Belief and Value, in Heuristics and Biases: The Psychology of Intuitive Judgment 120, 137 (Thomas Gilovich et al. eds., 2002).

[13] Gregory Mitchell, Why Law and Economics' Perfect Rationality Should Not Be Traded for Behavioral Law and Economics' Equal Incompetence, 91 Geo. L.J. 67, 133 & n.207 (2002).

The central idea of debiasing through substantive law is that in some cases it may be desirable to structure the substance of law – not merely the procedures by which the law is applied in an adjudicative setting – with an eye toward debiasing those who suffer from bounded rationality. As a simple intuitive example, consider "cooling-off" periods for consumer decisions.[14] The underlying concern is that under temporary pressures, consumers might make ill-considered or improvident decisions, at least in part resulting from bounded rationality. Responding to this concern, the Federal Trade Commission imposes a mandatory cooling-off period for door-to-door sales.[15] Under the Commission's rule, door-to-door sales must be accompanied by written statements informing buyers of their right to rescind purchases within three days of transactions.[16] (Some states also impose mandatory waiting periods before people may receive a divorce decree.[17]) Aware that people might act in a way that they will regret, regulators do not block their choices, but instead ensure a period for sober reflection. Thus, the law seeks to help people move in more rational directions – in this example through the simple and limited expedient of requiring a meaningful time window before decision making can occur.

As we have suggested, debiasing through substantive law targets actors out in the world, not participants (litigants or adjudicators) in the legal process. Part II below develops a series of examples of debiasing through substantive law, in areas ranging from consumer safety law to employment law to property law. The basic notion in each instance is to use the substantive content of law to improve the rationality of behavior, instead of trying to structure legal rules and institutions around the assumed persistence of actors' boundedly rational judgments.

Both the more familiar category of debiasing through procedural rules and the newer category of debiasing through substantive law raise important but little addressed normative questions. Compared to the usual approach of "debiasing law," an important advantage of strategies for debiasing through law is that they aim to correct errors while still preserving individuals'

---

[14] See Colin Camerer, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue & Matthew Rabin, Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism," 151 U. Pa. L. Rev. 1211, 1241-42 (2003), for a valuable discussion.

[15] 16 C.F.R. § 429.1(a) (2003).

[16] Id.

[17] See, e.g., Cal. Fam. Code. § 2339(a); Conn. Gen. Stat. Ann. § 46b-67(a).

opportunity to make choices. Unlike attempts to insulate legal outcomes from boundedly rational judgments, debiasing through law preserves a space for citizens to arrange their affairs as they like. Under the Commission's door-to-door sales rule, for example, no transaction is blocked. An important corollary is that, unlike "debiasing law" strategies, the approach of debiasing through law will frequently make it possible for government to improve outcomes for individuals who exhibit bounded rationality while leaving unrestricted the choices of those who would not otherwise err. It is preferable, when possible, to develop legal strategies that avoid imposing significant costs on those who do not exhibit boundedly rational behavior[18]; below we describe specific strategies that achieve this goal. In this important sense, debiasing through law provides real advantages over "debiasing law" strategies.

Still, at bottom, debiasing through law in either of its two varieties (substantive or procedural) involves the government in a self-conscious process of altering the behavior of at least some people by manipulating their perceptions of the reality around them. Such an approach raises obvious problems that require discussion. It would be possible to question any effort to use law to work against people's values and perceptions – especially, perhaps, if that approach is based on a sophisticated understanding of how values and perceptions can be altered. However, as we show, even the most uncontroversial provisions of the civil and criminal law are, in a sense, a form of "debiasing," designed to alter both values and perceptions. Nonetheless, there are legitimate objections to certain kinds of debiasing through law, and we trace some of those objections here, especially those involving the risk of overshooting and the danger of manipulation.

Part I below offers some preliminary remarks on the background for our analysis, relevant definitions, and the domain of our account. Part II, the heart of the Article, describes and analyzes a range of examples of debiasing through substantive law. Part III considers normative questions raised by debiasing through law.

I. Definitions and the Domain of Analysis

[18] See Camerer, Issacharoff, Loewenstein, O'Donoghue & Rabin, supra note 14, at 1212; Mitchell, supra note 13, at 132.

If debiasing through law is a response to bounded rationality, an obvious first step is to understand the basic idea of bounded rationality. Section A below offers an introductory discussion. The next threshold question for our analysis is what it means to debias boundedly rational actors. Existing treatments have typically focused on debiasing of individuals suffering from a discrete form of bounded rationality rather than on debiasing of boundedly rational individuals as a general matter.[19] The specificity of most of the existing literature has, among other things, deflected attention from foundational questions about how debiasing of boundedly rational actors should be defined. In section B we suggest the importance of distinguishing such debiasing from incentives, another instrument for affecting people's behavior. In section C we offer some general comments on the domain of our analysis of debiasing through law.

A. Some Notations on Bounded Rationality

What is the nature of the problems that debiasing through law is meant to address? As is now well-known, psychologists and behavioral economists have uncovered a wide range of departures from unboundedly rational behavior. These departures take one of two general forms. First, individuals may exhibit judgment biases, typically errors in forecasting the likelihood of events.[20] Second, human behavior may deviate from the precepts of expected utility theory.[21] We consider these two basic categories in turn.

Many of the most well-established judgment biases reflect the use of "heuristics." Heuristics are mental short-cuts, or rules of thumb, that generally work well but that produce systematic errors in some settings.[22]

---

[19] See, e.g., Sanna, Schwarz & Stocker, supra note 6. Fischhoff, supra note 6, is an exception.

[20] See generally Heuristics and Biases: The Psychology of Intuitive Judgment (Thomas Gilovich et al. eds., 2002).

[21] See generally Choices, Values, and Frames (Daniel Kahneman & Amos Tversky eds., 2001).

[22] For general discussions, see Daniel Kahneman & Shane Frederick, Representativeness Revisited, in Heuristics and Biases: The Psychology of Intuitive Judgment 49 (Thomas Gilovich et al. eds., 2002); Peter M. Todd, Fast and Frugal Heuristics for Environmentally Bounded Minds, in Bounded

Heuristics typically work through a process of "attribute substitution," in which people answer a hard question by substituting an easier one.[23] For example, people might resolve a question of probability not by investigating statistics, but by asking whether a relevant incident comes easily to mind.[24] The resulting "availability heuristic" often produces sensible judgments and behavior for people who lack detailed statistical information, but it can lead to significant and severe errors.[25] Similarly, the use of heuristics has been shown to lead people to misestimate probabilities by ignoring sample size and by committing the conjunction fallacy (concluding that characteristics X and Y are more likely to be present than characteristic X alone) – errors produced by the so-called representativeness heuristic.[26] (We will give specific examples of these forms of behavior below.[27]) Heuristics, then, are not themselves biases, but they can produce biases. Thus "availability bias" might be said to arise when the availability heuristic leads people to make predictable errors in assessing probabilities.

A related set of findings emphasizes not mental short-cuts, but more direct biases that leads to inaccurate judgments. We have already referred to three examples: optimism bias, hindsight bias, and self-serving bias.

We are also concerned with biases that are not strictly part of the heuristics-and-biases literature, but that can be understood in related terms, and that similarly lead to systematic errors. Unconscious biases based on race, sex, or disability, for example, will often lead employers to treat African-Americans, women, and individuals with disabilities less well than other employees. In a way, the characteristic of race, sex, or disability often operates as a heuristic. Attribute substitution is important here; the relevant

---

Rationality: The Adaptive Toolbox 51 (Gerd Gigerenzer & Reinhard Selten eds., 2002).

[23] See Kahneman & Frederick, supra note 22, at 53.

[24] See Norbert Schwartz & Leigh Ann Laughn, The Availability Heuristic Revisited: Ease of Recall and Content of Recall As Distinct Sources of Information, in Heuristics and Biases: The Psychology of Intuitive Judgment 103, 103 (Thomas Gilovich et al. eds., 2002).

[25] See Jolls, Sunstein & Thaler, supra note 1, at 1477-78.

[26] Amos Tversky & Daniel Kahneman, Judgments of and by Representativeness, in Judgment Under Uncertainty: Heuristics and Biases 84, 92-94 (Daniel Kahneman et al eds., 1982).

[27] See infra notes 39-40 and accompanying text (representativeness heuristic); infra note 85 and accompanying text (availability heuristic).

characteristic substitutes for a more fine-grained inquiry into relevant qualifications. Whether the heuristic is generally accurate (as in the case of rational statistical discrimination[28]) or not, it can produce severe and systematic mistakes in particular cases. The antidiscrimination laws forbid the conscious use of the relevant characteristics whether or not they are frequently accurate; employers are required to individuate even if individuation is costly.[29] An important remaining problem is the effect of *unconscious* bias, which the law has not addressed so clearly.[30]

The second broad category of bounded rationality consists of departures from expected utility theory.[31] Actual choices diverge in important ways from the predictions of expected utility theory. A leading alternative to that theory is prospect theory, according to which people evaluate outcomes based on the change they represent from an initial reference point, rather than based on the nature of the outcome itself; they also weigh losses more heavily than gains (and thus show "loss aversion").[32] An important implication of the loss aversion posited by prospect theory is the "endowment effect," according to which an individual's valuation of an entitlement depends on whether the individual is given initial ownership of that entitlement.[33] Thus, for example, individuals endowed with university mugs demand substantially more to sell these mugs than unendowed individuals are willing to pay to buy such mugs.[34] While this aspect of bounded rationality has received only modest

---

[28] See generally Kenneth J. Arrow, The Theory of Discrimination, in Discrimination in Labor Markets 3 (Orley Ashenfelter & Albert Rees eds., 1973); George Akerlof, The Economics of Caste and of the Rat Race and Other Woeful Tales, 90 Q.J. Econ. 599 (1976).

[29] See David A. Strauss, The Law and Economics of Racial Discrimination in Employment: The Case For Numerical Standards, 79 Geo. L.J. 1619, 1623 & n. 13 (1991).

[30] See Linda Hamilton Krieger, The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity, 47 Stan. L. Rev. 1161, 1166-86 (1995).

[31] See generally Choices, Values and Frames, supra note 21; Daniel Kahneman & Amos Tversky, Prospect Theory: An Analysis of Decision Under Risk, 47 Econometrica 263 (1979).

[32] Kahneman & Tversky, supra note 31, at 273.

[33] See Daniel Kahneman, Jack Knetsch & Richard Thaler, Experimental Tests of the Endowment Effect and the Coase Theorem, 98 J. Pol. Econ. 1325, 1325-26 (1990).

[34] See id. at 1329-42.

attention in the existing social science literature on debiasing of boundedly rational actors, Part II below suggests its relevance to debiasing through law.

B. Debiasing Versus Incentives

A number of different channels exist through which boundedly rational behavior may be made to "go away" or diminish in degree. Consider these examples:

(1) People are prone to social influences, so much so that many people will ignore the clear evidence of their own senses, and hence provide incorrect answers, if they are confronted with the unanimous views of others.[35] This kind of "conformity bias," in which the views of others are used as a kind of heuristic for the proper answer, is significantly reduced when financial incentives are provided. When people stand to gain economically from a correct answer and when they have confidence in their own judgment, they are far more likely to ignore the crowd, to say what they think, and to answer correctly.[36] So too, economic incentives have been found to reduce the errors associated with the cognitive bias of ignoring sample size.[37]

(2) Individuals in the role of litigants, having previously exhibited a tendency to see cases only in the light most favorable to their own side, are required to consider weaknesses in their side or reasons that the judge might rule against them. As noted earlier, the "self-serving bias" bias they had previously exhibited vanishes.[38]

(3) After reading a paragraph about a thirty-one year old woman, Linda, who was concerned with issues of social justice and discrimination in college, most people tend to say that Linda is more likely to be "a bank teller

---

[35] See Solomon Asch, Opinions and Social Pressure, in Readings About the Social Animal 13 (Elliott Aronson ed., 1995).
[36] See Robert Baron et al., The Forgotten Variable in Conformity Research: Impact of Task Importance on Social Influence, 71 J. Personality & Social Psych. 915 (1996).
[37] See Vidya Awashti & Jamie Pratt, The Effects of Monetary Incentives on Effort and Decision Performance: The Role of Cognitive Characteristics, 65 Accounting Rev. 797 (1990).
[38] Babcock, Loewenstein, & Issacharoff, supra note 8, at 918-19.

and active in the feminist movement" than to be "a bank teller."[39] This is an example of the conjunction fallacy, produced by the representativeness heuristic. But people are less likely to commit the conjunction fallacy when asked about frequencies rather than probabilities. If asked, "of 100 people who fit the description" of Linda, how many are bank tellers and how many are bank tellers and active in the feminist movement, the level of conjunction violations drops from 80% or more to 20% or less.[40]

(4) A group of individuals views pictures of Michael Jordan and Tiger Woods before submitting to testing of unconscious racial bias. Both immediately after viewing the pictures and twenty-four hours later, this group exhibits substantially less unconscious racial bias than individuals not exposed to the pictures of Jordan and Woods.[41]

The first of these four examples is one in which the boundedly rational behavior is eliminated by the provision of incentives. A broad definition of debiasing of boundedly rational actors might embrace this sort of technique, but we think it is preferable to exclude the underlying form of behavior here from the category of boundedly rational behavior (so that the removal of the behavior by the provision of incentives does not count as "debiasing" of boundedly rational actors in the sense that we understand that term). For some purposes, it might be useful to understand incentives as a way of overcoming boundedly rational behavior by increasing the stakes. Baruch Fischhoff, for instance, describes "rais[ing] stakes" as a possible strategy for debiasing of boundedly rational actors.[42] But it seems most conservative, and most consistent with existing conventions in analyses of bounded rationality, to limit the category of boundedly rational behavior to that which survives even in the presence of financial or other consequences for exhibiting the behavior.[43] If an apparent departure from unbounded

---

[39] Tversky & Kahneman, supra note 26, at 92-94.
[40] Gerd Gigerenzer, Adaptive Thinking 250 (2000).
[41] Nilanjana Dasgupta & Anthony G. Greenwald, On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals, 81 J. Pers. & Soc. Psych. 800, 803-04 (2001).
[42] Baruch Fischhoff, Heuristics and Biases in Application, in Heuristics and Biases: The Psychology of Intuitive Judgment 730, 732 (Thomas Gilovich et al. eds., 2002); Fischhoff, supra note 6, at 424 & tbl.1.
[43] Colin Camerer & Robin Hogarth, The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework, 19 J.

rationality is eliminated with the provision of financial incentives, then many would conclude that it was not a departure from unbounded rationality at all, but instead a mere result of lazy or careless decision making by an actor who had no reason to be other than lazy or careless. Under our approach, therefore, the technique used in the first example above is not a strategy for debiasing of boundedly rational actors. And the same goes for techniques that eliminate boundedly rational behavior by improving a previously faulty aspect of an experimental design – although here again Fischhoff's broad conception of debiasing of boundedly rational actors embraces such manipulations.[44]

The second example above is a standard case of debiasing of individuals exhibiting bounded rationality. Subjects are asked to consider arguments of a particular sort, and the consideration of such arguments eliminates the boundedly rational behavior they previously exhibited. Importantly, the technique here differs from incentives. Agents are not asked to repeat the very same task with the very same structure, with the sole difference that they now have greater reason to take care in making their choices; instead the environment is restructured in a way that alters not their motivation but the actual process by which they perceive the reality around them. Thus, we define debiasing of boundedly rational actors as using techniques that intervene in and alter the situation that produces the boundedly rational behavior, without operating on the degree of motivation or effort a subject brings to the task.

The third case, involving the conjunction fallacy, is a simple illustration of debiasing through reframing. As we shall see, the harmful effect of some heuristics can be reduced, and some biases can be eliminated, either through reframing or through the behaviorally informed presentation of information. In many domains, debiasing of boundedly rational actors will not occur through the provision of information alone (itself a naive notion, since some mode of presentation is inevitable, as discussed in some detail below). Steps must be taken to ensure that the information is presented in such a way as to address the bias. We discuss normative dimensions of this idea below.

Risk & Uncertainty 7 (1999), make this implicit claim, and offer a great deal of evidence that many cases of boundedly rational behavior are not eliminated by the provision of incentives.

[44] Fischhoff, supra note 6, at 424 & Table 1.

The fourth case, concerning the manipulation of unconscious racial bias, is also an example of debiasing of individuals who previously exhibited biased behavior. Exposure to the pictures of Jordan and Woods does not directly affect the degree of motivation or effort a subject will bring to bear on the assigned task, but it produces significant changes in the degree of unconscious racial bias. It has sometimes been suggested in the law review literature that unconscious bias "cannot be reliably manipulated or controlled"[45]; but a substantial recent literature in social science suggests significant malleability of unconscious bias. Thus debiasing of racial and other unconscious biases represents a potentially important avenue for law – one that we explore at some length below.

The idea of debiasing of boundedly rational actors can be connected to "dual process" approaches of the sort that have received considerable recent attention in psychology.[46] According to such approaches, people use two cognitive systems. System I is rapid, intuitive, and error-prone; System II is more deliberative, calculative, slower, and more likely to be error-free. Heuristic-based thinking is rooted in System I; it is subject to override, under certain conditions, by System II.[47] People can and do use their own System II to correct the blunders produced by System II. We might think of debiasing strategies of the sorts emphasized in this Article as an effort to activate System II in order to reduce the risk of mistake. As we shall see, however, some of these debiasing strategies actually attempt to enlist System I, correcting mistakes by invoking heuristics themselves.

C. The Domain of Analysis

---

[45] Amy L. Wax, Discrimination as Accident, 74 Ind. L.J. 1129, 1158 (1999); see also Krieger, supra note 30, at 1245.

[46] See generally Social Judgments (Joseph P. Forgas et al. eds., 2003); Shelly Chaiken & Yaacov Trope, Dual-Process Theories in Social Psychology (1999); Kahneman & Frederick, supra note.

[47] See Kahneman & Frederick, supra note 22, at 51. The two systems need not be seen as occupying different physical spaces; they might even be understood as heuristics (!), see id. There is, however, some evidence that different sectors of the brain can be associated with Systems I and II. See Arne Ohman & Stefan Wiens, The Concept of An Evolved Fear Module and Cognitive Theories of Anxiety, in Feelings and Emotions: The Amsterdam Symposium 58 (Antony S.R. Manstead et al. eds. 2004); Joseph LeDoux, The Emotional Brain 106-132 (1996).

With any aspect of bounded rationality – whether a judgment bias or a departure from expected utility theory – it is always possible that the behavior in question is offset to some degree by another aspect of bounded rationality that tends in the opposite direction. Some biases can correct others. In such cases, debiasing efforts directed to one aspect of bounded rationality might actually make things worse rather than better – a clear application of the theory of second best.[48] Whether a given aspect of bounded rationality is in fact likely to be in an offsetting relationship with some other feature of human behavior will obviously depend on the particular context.[49] We focus below on situations in which there is no readily apparent counterforce to the departure from bounded rationality that argues for debiasing through law.

It is also important to stress that not all departures from unbounded rationality seem to respond well to techniques for debiasing of boundedly rational actors. Consider, for instance, hindsight bias. Not only do many manipulations fail to reduce this bias, but some seemingly sensible strategies for debiasing actors who exhibited the bias have actually increased it.[50] To be sure, studies that have required subjects to "rethink the inferences that they have made upon learning [an] outcome and [have then] demonstrated to them that other inferences remained plausible" have shown some success in reducing hindsight bias.[51] Indeed, in one study, a defense counsel who warned mock jurors not to be "Monday-morning quarterbacks" and to avoid second guessing the decisions of the defendant's reduced hindsight bias in these jurors by over seventy percent.[52] However, in most cases strategies – even fairly aggressive ones – for debiasing boundedly rational actors have enjoyed limited, if any, success in combating hindsight bias.[53]

---

[48] See Gregory Besharov, Second-Best Considerations in Connecting Cognitive Biases (working paper).

[49] See, e.g., Jolls, Sunstein & Thaler, supra note 1, at 1524 (discussing the partially offsetting relationship between hindsight bias and optimism bias in the tort law context).

[50] See Sanna, Schwarz & Stocker, supra note 6.

[51] Rachlinski, supra note 4, at 586-88.

[52] Merrie Jo Stallard & Debra L. Worthington, Reducing the Hindsight Bias Utilizing Attorney Closing Arguments, 22 J. L. & Hum. Behav. 671, 680-81 (1998).

[53] See Fischhoff, supra note 6, at 427-31; Reid Hastie & W. Kip Viscusi, What Juries Can't Do Well: The Jury's Performance as a Risk Manager, 40

But in other contexts techniques for debiasing boundedly rational actors have shown substantial promise. One example is the case of self-serving bias discussed above; as noted, having subjects consider the weaknesses in their case or reasons that the judge might rule against them appears effective in eliminating self-serving bias.[54] A second example, also mentioned above, is the way in which unconscious bias against members of particular groups may be reduced upon exposure to bias-challenging stimuli.[55] A third example involves the neglect of salient variables in the domain of risk, such as probability of harm and numbers of people in jeopardy. When steps are taken to draw people's attention to these variables, they are far less likely to be neglected.[56]

The domain of our analysis in this Article will be those cases in which debiasing of boundedly rational actors has shown a strong likelihood of success in the existing social science literature. Within this domain, we are interested in assessing the degree to which the law can, does, and should play a role in achieving such debiasing. (Thus our analysis bears but does not focus on forms of debiasing – such as Robert Rasmussen's example of credit scoring, thought to check optimism bias by loan officers[57] – that might be undertaken by private actors independent of direct legal intervention.) As noted above, we focus particularly on debiasing through substantive law, the subject of Part II below.

Ariz. L. Rev. 901, 917 (1998); Kim A. Kamin & Jeffrey J. Rachlinski, Ex Post ≠ Ex Ante: Determining Liability in Hindsight, 19 L. & Hum. Behav. 89, 97-98 (1995); Jeffrey J. Rachlinski, Regulating in Foresight Versus Judging Liability in Hindsight: The Case of Tobacco, 33 Ga. L. Rev. 813, 824 (1999).

[54] Babcock, Loewenstein & Issacharoff, supra note 8, at 918-19.

[55] Dasgupta & Greenwald, supra note 41, at 803-04.

[56] See Chris Hsee, The Evaluability Hypothesis: An Explanation for Preference Reversals Between Joint and Separate Evaluations of Alternatives, 67 Org. Behav. & Human Decision Process 242 (1996). See generally Howard Margolis, Dealing With Risk (1996).

[57] Robert K. Rasmussen, Behavioral Economics, The Economic Analysis of Bankruptcy Law and the Pricing of Credit, 51 Vand. L. Rev. 1679, 1695 (1998).

II. Debiasing Through Substantive Law

Figure 1 maps the terrain of debiasing through law a bit more fully. The column division marks the line between procedural rules governing the adjudicative process and substantive rules regulating actions taken outside of the adjudicative process. The row division marks the line between debiasing actors in their capacity as participants in the adjudicative process and debiasing actors in their capacity as decision makers outside of the adjudicative process. The upper left box in this matrix represents the type of debiasing through law on which the existing literature has focused: the rules in question are procedural rules governing the adjudicative process, and the actors targeted are individuals in their capacity as participants in the adjudicative process. (Of course, the simple fact that lawyers play a role in the adjudicative process may blunt the effects of litigants' or even adjudicators' bounded rationality[58]; nonetheless we believe that strategies for debiasing through procedural rules hold important promise, though they are not our focus in this Article.)

Moving counterclockwise, the lower left box in the matrix is marked with an "X" because procedural rules governing the adjudicative process do not have any obvious role in debiasing actors outside of the adjudicative process – although they certainly may affect such actors' behavior in various ways by influencing what would happen in the event of future litigation. (Perhaps there are even feasible mechanisms for debiasing, through the change of procedural rules governing the adjudicative process, actors in their capacity as decision makers outside of the adjudicative process; but we do not explore the possibility here.) The lower right box in the matrix represents the category of debiasing through law emphasized in this Article: the rules in question are substantive rules regulating actions taken outside of the adjudicative process, and the actors targeted are decision makers outside of the adjudicative process.[59]

---

[58] See generally Russell Korobkin & Chris Guthrie, Psychology, Economics, and Settlement: A New Look at the Role of the Lawyer, 76 Tex. L. Rev. 77 (1997).

[59] As will become clear, we understand "substantive law" in a broad sense, to include initiatives that are, in one sense, procedural, such as information campaigns. But we exclude from our category of "substantive law" rules regulating the adjudicative process; these rules are the focus of existing work on debiasing through law in the legal literature.

Figure 1: Typology of Strategies for Debiasing Through Law

	Type of Law		
		Procedural rules governing the adjudicative process	Substantive rules regulating actions taken outside of the adjudicative process
Role of Actor	Debiasing actors in their capacity as participants in the adjudicative process	Debiasing through procedural rules	"Hybrid" debiasing
	Debiasing actors in their capacity as decision makers outside of the adjudicative process		Debiasing through substantive law

Finally, the upper right corner of the matrix represents a hybrid category that warrants brief discussion, in part to demarcate it from the category on which we focus in this Article. In this hybrid category, it is substantive, rather than procedural, law that is manipulated to achieve debiasing, but the aspect of bounded rationality that the debiasing effort targets is one that arises within, rather than outside of, the adjudicative process. For example, Ward Farnsworth's recent work on self-serving bias suggests that such bias on the part of employment discrimination litigants (actors in their capacity as participants in an adjudicative process) might be reduced by restructuring employment discrimination standards (substantive rules regulating action outside of the adjudicative process) to increase the reliance of such standards on objective facts as opposed to subjective or normative judgments.[60] Farnsworth's debiasing suggestion operates through reform of substantive rather than procedural law, but the actions to be debiased are those of litigants within the adjudicative process.

[60] Farnsworth, supra note 11, at 593-95.

Our treatment below of debiasing through substantive law explores a number of examples, which fall into four basic categories. These are debiasing through restrictions on practices and communications by firms (sections A and B below); debiasing through the structure of certain legal-organizational forms and remedies and forms (sections D and F below); debiasing through switching the default rule (section E below); and (a category that may already be familiar to some) debiasing through government provision of information (section C below). In each case we focus on situations in which the departure from unbounded rationality that is the target of the attempt at debiasing through law is itself a robust feature of the behavior of a substantial number of individuals, and in which this departure from unbounded rationality is not likely to be offset by other, distinct aspects of bounded rationality.

A. Debiasing Through Consumer Safety Law

Many federal and state laws regulate the safety of products used by consumers.[61] A central impetus for these laws is the idea that consumers often do not adequately perceive the potential risks of such products. This section describes a leading reason that consumers may not adequately perceive such risks, and then discusses ways that the law might attempt to debias such individuals through consumer safety law. Such an analysis suggests the potential virtues of debiasing through law strategies as a complement to or substitute for the conventional "debiasing law" strategies generally employed in the consumer safety area.[62] Our analysis shares a starting point with existing proposals for better informing consumers,[63] but comes to a quite different end point given our behavioral appreciation of the limits of some forms of information provision; thus our hope is that this example, with its overlapping relationship to some existing proposals in the consumer safety area, is a good initial lens on the central attributes of our conception of debiasing through substantive law.

1. Optimism Bias and Consumer Safety

[61] See, e.g., 15 U.S.C. §§ 2601-2692 (2000) (Toxic Substances Control Act); 15 U.S.C. §§ 2051-2084 (2000) (Consumer Product Safety Act).
[62] See sources cited supra note 61.
[63] See, e.g., Joseph E. Stiglitz, Economics of the Public Sector 90-91 (1986).

A common feature of human behavior is optimism bias; most people tend to think their probability of a bad outcome is far less than others' probability – although of course this cannot be true for everyone. People typically think that their chances of having an auto accident, contracting a particular disease, or getting fired from a job are significantly lower than the average person's chances of suffering these misfortunes.[64] Estimates offered by individuals for their own probabilities range from twenty to eighty percent below the average person's probability.[65]

While the "above average" effect is well-established, it does not establish that people optimistically underestimate their statistical risk.[66] People could believe, for example, that they are less likely than most people to contract cancer, while also having an accurate sense of the probability that they will contract cancer; this would be because they overestimate the probability that *others* will contract cancer. But substantial evidence suggests that people sometimes exhibit optimism bias in the estimation of actual probabilities, not simply relative risk. For example, professional financial experts consistently overestimate likely earnings, and business school students overestimate their likely starting salary and the number of offers that they will receive.[67] People also underestimate their own likelihood of being involved in a serious automobile accident,[68] and their frequent failure to buy insurance for floods and earthquakes is consistent with the view that people are excessively optimistic.[69]

It bears noting as well that these data pointing to optimism bias come from individuals making judgments that they make regularly in their everyday lives, rather than judgments far removed from those they would ordinarily make.[70] But as Daniel Armor and Shelley Taylor have emphasized,

---

[64] See Christine Jolls, Behavioral Economic Analysis of Redistributive Legal Rules, 51 Vand. L. Rev. 1653, 1659-62 (1998) (discussing studies).
[65] Id. at 1659.
[66] See W. Kip Viscusi, Smoke-Filled Rooms 162-66 (2002).
[67] See David Armour & Shelley E. Taylor, When Predictions Fail: The Dilemma of Unrealistic Optimism, in Heuristics and Biases: The Psychology of Intuitive Judgment 334, 334-35 (Thomas Gilovich et al. eds. 2002).
[68] See Jolls, supra note 64, at 1660-61.
[69] See id. at 1661.
[70] Cf. Chris Guthrie, Prospect Theory, Risk Preference, and the Law, 97 NW. U. L. Rev. 1115, 1156-57 (2003) (discussing relevance of study designs in

optimism bias is context-dependent.[71] The extent of the bias decreases when people are in the midst of deliberating, when the outcome will be known in the near future, and when the consequences of error are especially severe.[72] Nevertheless, the evidence of bias is sufficient to suggest that many people do fall prey to it.

In the context of consumer safety, consumers may not adequately perceive product risks because they are imperfectly informed, because they are optimistically biased, or both. The traditional economic view is that the problem (if there is one at all) is merely a problem of imperfect information, and thus is appropriately corrected by provision of additional information.[73] In a world of bounded rationality, however, the prescription of "more information" is naïve and incomplete.[74] Among other things, presentation of the information greatly matters, and any presenter must make choices about framing. In addition, optimism bias will lead many people to underestimate their personal risks even if they accurately understand average risks.[75]

It would be reasonable to conclude that optimism bias justifies heightened standards of products liability as an alternative to the provision of additional statistical facts about the product in question. Jon Hanson and Douglas Kysar, for instance, argue in favor of enterprise liability on the basis of boundedly rational behavior such as optimism bias.[76] However, such an approach – seeking to "debias law" – is thought by other scholars to impose large costs of its own.[77] A still more aggressive approach, available in the case of some products, is an across-the-board ban on the product's use. A number of federal statutes give agencies a choice among disclosure requirements and partial or complete bans.[78] In response to evidence of inadequate information, optimism, and other consumer biases, some

---

which subjects were making decisions of a type that they were accustomed to making).

[71] Armour & Taylor, supra note 67, at 338-41.

[72] Id.

[73] See, e.g., Stiglitz, supra note 63, at 90-91.

[74] Jolls, Sunstein & Thaler, supra note 1, at 1542.

[75] See, e.g., Hanson & Kysar, supra note 2, at 729-30; Dave Slovic, Do Adolescent Smokers Know the Risks?, 47 Duke L.J. 1133, 1137 (1998).

[76] Hanson & Kysar, supra note 3, at 1560.

[77] See, e.g., George L. Priest, The Current Insurance Crisis and Modern Tort Law, 96 Yale L.J. 1521, 1560 (1987)

[78] See sources cited supra note 61.

regulators might well be tempted to impose a ban even if the statute reflects a preference for disclosure.[79]

An important and largely unexplored alternative to these "debiasing law" strategies is to use the law to reduce the occurrence of boundedly rational behavior. At the broadest level, strategies for debiasing through consumer safety law provide a sort of middle ground between inaction or the economists' spare prescription of "more information," on the one hand, and the aggressive "debiasing law" strategies of heightened products liability standards or outright bans, on the other. Strategies for debiasing through law, alert to the source of the underlying problems faced by consumers, may be far more successful than a less informed informational strategy, and also far more protective of consumer prerogatives than the strategy of an across-the-board ban. Below, we outline two possibilities for debiasing through law in the consumer safety area. As will become clear, an important theme of our discussion is that effective strategies for debiasing through law may themselves harness separate departures from unbounded rationality – a feature of debiasing through law that raises normative issues explored further in Part III below.

In the discussion to follow, we will focus on the scenario in which optimism bias is likely to produce an *overall* underestimation by consumers of the risk associated with a given product. As we have already suggested in general terms, a competing departure from unbounded rationality could in some circumstances lead consumers to overestimate rather than underestimate the risk associated with the product. For instance, highly available instances of accident or injury can lead to excessive pessimism -- a distortion opposite to the one produced by optimism bias.[80] Unless, however, there is a reason to think that events of accident or injury associated with a given product are highly available – a classic example here is a plane crash[81]

---

[79] See Corrosion Proof Fittings v. EPA, 947 F.2d 1201 (5th Cir. 1991) (interpreting Toxic Substances Control Act to require the least restrictive regulatory alternative).

[80] See Alan Schwartz & Louis L. Wilde, Imperfect Information in Markets for Contract Terms: The Examples of Warranties and Security Interests, 69 Va. L. Rev. 1387, 1437 (1983).

[81] See Cass R. Sunstein, Bad Deaths, 14 J. Risk & Uncertainty 259, 259 (1997).

– optimism bias suggests that many consumers will tend toward underestimation of the risks associated with products they use.[82]

2. Debiasing Through the Availability Heuristic

In response to the risk that optimistically biased people believe "it won't happen to them," one might imagine adopting strategies such as suggesting reasons that negative outcomes might occur or considering risk factors related to negative outcomes. However, such approaches have usually failed to reduce optimism bias.[83] Successful strategies for debiasing through law in the consumer safety context may well require harnessing other aspects of boundedly rational behavior.[84]

Consider, as a possible response to optimism bias, the availability heuristic described earlier. Here is a familiar example of availability: Individuals asked how many words in a 2,000-word novel end in "ing" give much larger estimates than individuals asked how many words have "n" as the second-to-last letter, notwithstanding the obvious fact that more words satisfy the latter criterion than the former.[85] Recall here that heuristics typically operate through a process of attribute substitution; in this light, the basic idea of availability is that individuals attach higher likelihoods to events or outcomes that they have greater ease calling to mind. Use of the availability heuristic often produces a form of judgment error. As with optimism bias, availability can lead to systematic mistakes in the assessment of probabilities. (Thus "availability bias," in the form of excessively high estimates, and "unavailability bias," in the form of excessively low estimates, involve complementary errors stemming from the use of this heuristic.) But because making an occurrence "available" can cancel out the direction of the

---

[82] See, e.g., Hanson & Kysar, supra note 2, at 729-30.

[83] See Weinstein & Klein, supra note 6, at 322-23.

[84] See Steven J. Sherman, Robert B. Cialdini, Donna F. Schwartzman, & Kim D. Reynolds, Imagining Can Heighten or Lower the Perceived Likelihood of Contracting a Disease: The Mediating Effect of Ease of Imagery, in Heuristics and Biases: The Psychology of Intuitive Judgment 98, 98 (Thomas Gilovich et al. eds., 2002).

[85] Amos Tversky & Daniel Kahneman, Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment, 90 Psychol. Rev. 293, 295 (1983).

optimism bias, availability holds clear promise as a strategy for debiasing of boundedly rational actors.[86]

Consider, for instance, the empirical findings of Neil Weinstein suggesting that many people substantially underestimate their risk of cancer.[87] If we imagine that women asked to estimate their risk of breast cancer are also told a poignant and detailed story about a woman their age with similar family and other circumstances who was diagnosed with breast cancer, then their estimated probabilities are likely to be higher (although they may of course be too much higher or not enough higher – points to which we return in Part III below). The point here is similar in spirit to Chris Guthrie's suggestion that legal policy makers bring "vivid information about plaintiff losses in frivolous litigation" to bear in reducing plaintiffs' overestimation of the probability of success in such litigation.[88] In fact a recent study of smoking behavior finds a phenomenon of exactly this kind: many smokers are unrealistically optimistic, but their judgments become more realistic when presented with vivid evidence of the health harms that accompany smoking.[89] In the absence of such evidence, smokers often lack a real sense of harms to quality of life from smoking, but presented with such evidence they are better informed and often alter their behavior.[90] Here is a successful effort at debiasing by enlisting availability. The problem of obesity, now subject to a range of legal initiatives,[91] could easily be approached in this way. If obesity is in part a product of optimism about associated health risks, along with self-control problems, then debiasing through the availability heuristic would be a natural response.

In the context of consumer safety law, debiasing through the availability heuristic would focus on putting at consumers' cognitive disposal the prospect of negative outcomes from use, or at least unsafe use, of a particular product. Specifically, the law could impose specific requirements

---

[86] Sherman, Cialdini, Schwartzman & Reynolds, supra note 84, at 259.
[87] Neil O. Weinstein, Unrealistic Optimism About Future Life Events, 39 J. Personality & Soc. Psychol. 806, 810 tbl.1 (1980).
[88] Chris Guthrie, Framing Frivolous Litigation: A Psychological Theory. 67 U. Chi. L. Rev. 163, 210 (2000).
[89] Frank A. Sloan, Donald H. Taylor, & V. Kerry Smith, The Smoking Puzzle: Information, Risk Perception, and Choice 122-23, 127, 161 (2003).
[90] Id. at 180-81.
[91] See, e.g., Note, Living on the Fat of the Land, 81 Wash. U.L. Q. 859 (2003).

on the way that information about the product would be presented to consumers. Firms could be required – on pain of administrative penalties or tort liability – to provide a truthful account of consequences that resulted from a particular harm-producing use of the product, rather than simply providing a generalized warning that fails to harness availability. (It is possible that "debiasing law" strategies such as enterprise liability would give some firms indirect incentives to provide such accounts.[92]) To enhance the efficacy of this strategy for debiasing through law, the law could further require that the real-life story of accident or injury be printed in large type and displayed prominently, so that consumers would be likely to see and read it before using the product. Mandatory warnings could conceivably raise First Amendment issues, but so long as there is no political or ideological disagreement with the content of the message, such warnings are likely to be constitutional.[93] The available evidence suggests that the approach of requiring the specific account as opposed to the generalized warning would help to reduce optimistic bias.[94] This, of course, is simply a species of our earlier suggestion that the *way* information is provided may be just as important as (or more important than) *that* information is provided.

It bears noting that a debiasing effort harnessing availability to counteract optimism bias would improve not only the decision making of consumers suffering from optimism bias but also the decision making of consumers suffering from simple information failures. A conspicuous, prominent account of injury from a product may help to correct the estimated probability of harm attached to the product by an optimistically biased consumer; at the same time, it should improve the behavior of imperfectly informed but not necessarily biased consumers.

If consumer safety law were to exploit availability in this way, an important aspect of the effort to achieve debiasing through law would be the modesty of the effort's scope, along two separate dimensions. First, consumers might begin to suffer from "information overload" if every time they went to buy any product – from a lawnmower to a candy bar to a fast food hamburger – they were hit with a real-life story of an individual harmed by use or consumption of the product. Their natural response might be to tune out all of the accounts provided by firms, even assuming these accounts were

---

[92] See Douglas A. Kysar, The Expectations of Consumers, 103 Colum. L.Rev. 1700, 1786 n.364 (2003).
[93] Glickman v. Wileman Brothers & Elliott, 512 U.S. 1145 (1997).
[94] Sherman, Cialdini, Schwartzman & Reynolds, supra note 84, at 98.

prominently displayed.[95] A successful strategy for achieving debiasing through law would need to target a limited number of discrete products for which the optimism bias problem was especially severe.

Likewise, the law would need to avoid overreaching in the severity of the featured outcomes. Firms should not be required to provide anecdotes reflecting highly unusual consequences of using their products; only if an outcome occurs with some frequency should the law seek to induce firms to make consumers aware of the prospect. An emphasis on worst-case scenarios might produce excessive responses.[96] Of course there are line-drawing problems here (as there are under conventional "debiasing law" strategies), but the basic point is straightforward: if requirements of anecdote-based warnings sweep in extremely unusual or unlikely scenarios, it is plausible that consumers will overreact or lose faith and fail to attach any weight at all to the accounts.

Worst-case scenarios may be much more easily avoided with a legal requirement that firms provide truthful anecdotes about genuine harms than with the alternative strategy – frequently utilized by government – of public information campaigns concerning risky products. Such campaigns have often resulted in the use of extremely vivid and salient images, to the point of producing overreaction or even backlash as a result of citizens' perceptions of manipulation. In the smoking context, for instance, the European Union has experimented with requirements that a percentage of cigarette packages sold in Europe have their fronts covered with vivid pictures of rotting teeth and blackened lungs.[97] The Canadian Health Ministry has required not only clear warnings ("Cigarettes cause strokes," "Tobacco smoke hurts babies," "Don't poison us," and "Tobacco can make you impotent") but also graphic pictures such as bleeding gums and two lungs with cancerous tumors.[98] Similarly, in the United States a well-known anti-drug advertisement from the 1980s featured a picture of an egg frying in a pan with the voiceover, "This is your

---

[95] See, e.g., W. Kip Viscusi, Individual Rationality, Hazard Warnings, and the Foundations of Tort Law, 48 Rutgers L. Rev. 625, 665-66 (1996).
[96] See generally Cass R. Sunstein, Probability Neglect: Emotions, Worst Cases, and Law, 112 Yale L.J. 61 (2002).
[97] See http://www.eplp.org.uk/PRTOBAC210301.html; http://lists.essential.org/pipermail/intl-tobacco/2001q1/000426.html
[98] See http://www.abc.net.au/news/science/health/2000/12/item20001224133940_1.htm

brain on drugs."[99] The suggestion of requiring, on pain of administrative sanctions or tort liability, truthful narratives of harm is a more modest and measured response to consumer optimism bias than such alternative approaches, which harness availability by aggressively exploiting highly salient, gripping images and which for this very reason may run an especially high risk of manipulation, overshooting, and other problems.[100]

The idea of debiasing through the availability heuristic is not without analogies in both regulatory law and existing practice. In the context of cigarette advertising, for instance, companies in some countries have been required to make their warnings as specific as possible, signaling the risks that smokers are likely to face. As recent evidence suggests, this response makes good behavioral sense.[101] Likewise, in the United States the American Legacy Foundation, a non-profit organization founded out of the 1998 settlement agreements between the tobacco industry and state attorneys general, launched an information campaign employing a close parallel to the debiasing strategy described above. The Foundation has publicized parting letters to children and other loved ones from mothers dying of smoking-related diseases; for instance, one letter reads, "Dearest Jon, I am so sorry my smoking will cheat us out of 20 or 30 more years together. Remember the fun we had every year at the lake. I will ALWAYS love and treasure you. Linda."[102]

3. Debiasing Through Framing

As noted in Part I.A, departures from unbounded rationality may take the form either of judgment biases (such as optimism bias) or departures from expected utility theory. An important feature of the leading alternative to expected utility theory – Daniel Kahneman and Amos Tversky's prospect theory – is that in evaluating outcomes people tend to weigh losses more heavily than gains.[103] It follows that framing the presentation of information

---

[99] Shaila K. Dawan, The New Public Service Ad: Just Say "Deal with It," N.Y. Times, Jan. 11, 2004, §3, at 5.
[100] See W. Kip Viscusi, Smoking 61-86 (1992) (suggesting that individuals do not underestimate, and may well overestimate, the risks from tobacco).
[101] See Sloan, Taylor & Smith, supra note 89, at 180-81.
[102] See http://women.americanlegacy.org/index.cfm.
[103] Kahneman & Tversky, supra note 31, at 273.

to exploit the extra weight attached to losses[104] is an additional means of counteracting optimism bias in the consumer safety context.

Returning to the case of breast cancer noted above, a well-known illustration of the effects of framing is a study involving breast self-examination.[105] Material that describes the positive effects of breast self-examination – such as a higher chance of discovering a tumor at an earlier stage – is ineffective.[106] By contrast, significant behavioral changes result from material that stresses the negative consequences of failing to undertake self-examination – such as a decreased chance of discovering a tumor when it remains treatable.[107] A recent example of the real-world recognition of the importance of framing effects was the vigorous dispute over whether government advertisements promoting breastfeeding in the United States should refer to the risks of leukemia and other childhood diseases from not breastfeeding (the approach favored by breastfeeding advocates) or instead to the benefits from breastfeeding (the approach viewed as more tolerable by infant formula manufacturers).[108] Showing an intuitive understanding of prospect theory, the infant formula manufacturers preferred that government emphasize the benefits of breastfeeding over the affirmative harms of not breastfeeding.

In the consumer safety context, framing effects point toward potentially effective methods of debiasing through law. Simple requirements that firms – such as infant formula manufacturers – "provide information" may be ineffective in part because firms' interest will be in framing the information in a way that minimizes the risks perceived by consumers. By contrast, a legal requirement that firms identify the negative consequences associated with their product or with a particular use of their product, rather than the positive consequences associated with an alternative product or with an alternative use of their product, is likely to be a highly effective means of

---

[104] For a recent overview of framing and loss aversion, see Barbara A. Mellers, Pleasure, Utility, and Choice, in Feelings and Emotions: The Amsterdam Symposium 282 (Antony R. Manstead et al. eds., 2004).
[105] Beth E. Meyerowitz & Shelly Chaiken, The Effect of Message Framing on Breast Self-Examination: Attitudes, Intentions, and Behavior, 52 J. Personality & Soc. Psychol. 500 (1987).
[106] Id. at 505.
[107] Id.
[108] See Melody Petersen, Breastfeeding Ads Delayed By a Dispute Over Content, N.Y. Times, Dec. 4, 2003, at C1.

reducing optimism bias exhibited by consumers. Particularly when coupled with the earlier recommendation based on the availability heuristic, such a step could make significant progress toward ensuring that consumers have a more accurate understanding of the risks associated with particular products, and could reduce the need for either a complete ban on some of the products in question or other "debiasing law" solutions.

B. Debiasing Through Employment Discrimination Law

A central and vexing issue in modern antidiscrimination law is the problem of unconscious bias.[109] While the most obvious target of employment discrimination prohibitions, such as Title VII of the Civil Rights Act of 1964, is discrimination that is consciously motivational in nature, many acts of discrimination now result from unconscious bias, understood as group-based devaluation of which the discriminator is not consciously aware.[110] An employer might well harbor no racial "animus," and sincerely disclaim and reject prejudice, but nonetheless act in response to subtle biases that affect decisions about hiring, firing, and conditions of employment. It is possible to argue that unconscious bias is accurate or rational,[111] as undoubtedly it is in some contexts; but under existing law, employment discrimination that grows out of such bias is a clear wrong, and, accordingly, we will assume that reducing such bias is a valuable objective for antidiscrimination law.

Alert to the problem of unconscious bias, commentators have responded with a variety of suggested legal reforms to deal with this problem. These reforms include replacing the current pretext model of disparate treatment proof with a "motivating factor" analysis that looks to whether bias played *some* role in the decision maker's behavior; permitting a cause of action for "nonwillful" discrimination (although perhaps with lesser remedies than those available for "willful" discrimination); and adopting a negligence

---

[109] See, e.g., Charles R. Lawerence III, The Id, The Ego, and Equal Protection: Reckoning with Unconscious Racism, 39 Stan. L. Rev. 317 (1987); David Benjamin Oppenheimer, Negligent Discrimination, 141 U. Pa. L. Rev. 899 (1993)

[110] See, e.g., Krieger, supra note 30, at 1164; Oppenheimer, supra note 109, at 900-17; Susan Sturm, Second Generation Employment Discrimination: A Structural Approach, 101 Colum. L. Rev. 458, 460, 468-69 (2001).

[111] See Wax, supra note 45, at 1142-43.

approach to discrimination.[112] These proposed reforms, while worthy of consideration, are notable for taking the fact of unconscious bias entirely as a given, rather than as something itself to be examined and possibly reduced or even undone.[113] As Linda Hamilton Krieger expresses the assumption, "[W]e [don't] know enough about how to reduce cognition-based judgment errors to enable us to translate such a duty into workable legal rules [because c]ognitive psychologists have told us more about the shortcomings of human social inference cognition than about how the various biases they identify can be reduced or controlled."[114]

Our analysis here is focused on an alternative possibility – that both existing aspects and possible reforms of Title VII might attempt to strike at the underlying problem of unconscious bias by seeking directly to debias employment decision makers. We begin with some background about a leading method for measuring unconscious bias and then discuss prospects for debiasing through employment discrimination law. A central claim here, as elsewhere, is that some pockets of current law actually attempt to engage in such debiasing.

1. Evidence and Nature of Unconscious Bias

Scholars from a wide range of fields have identified diverse means of assessing and measuring unconscious bias,[115] and we do not begin to attempt a summary or comprehensive treatment here. We give primary emphasis to a leading technique from the modern social psychology literature for measuring unconscious bias; this technique is the Implicit Attitudes Test (IAT), which has had widespread influence in part because of its ready availability on the Internet and the resulting large number of individuals who have taken the test.

---

[112] See Krieger, supra note 30, at 1241-47; Oppenheimer, supra note 109, at 967-72.

[113] Krieger, supra note 30, at 1245.

[114] Id.; see also Wax, supra note 45, at 1133 (similar).

[115] See, e.g., Samuel L. Gaertner & John P. McLaughlin, Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics, 46 Soc. Psychol. Q. 23 (1983); Anthony G. Greenwald & Mazharin R. Banaji, Implicit Social Cognition: Attitudes, Self-Esteem and Stereotypes, 102 Psychol. Rev. 4 (1995); Lawrence, supra note 109.

In the IAT, respondents are asked to categorize a series of stimuli (words or pictures) into four groups, two of which are demographic categories (such as "black" and "white"), and the other two of which are the categories "good" and "bad." Groups are paired, so that a respondent would be asked to press one key on the computer for either "black" or "bad" and a different key for either "white" or "good" (a stereotype-consistent pairing); or would be asked to press one key on the computer for either "black" or "good" and a different key for either "white" or "bad" (a stereotype-consistent pairing). Stimuli are (for example) pictures of black faces, pictures of white faces, "good" words such as joy, love, peace, wonderful, pleasure, glorious, laughter, and happy, and "bad" words such as agony, terrible, horrible, nasty, evil, awful, and failure. Unconscious bias is defined as faster categorization when the "black" and "bad" categories are paired than when the "black" and "good" categories are paired.

The results of the IAT are striking. Three-quarters of respondents exhibit faster categorizations with the stereotype-consistent pairing (black-bad and white-good) than with the stereotype-inconsistent pairing (black-good and white-bad), while twelve percent exhibit no difference in speed of categorization. Only sixteen percent of respondents exhibit faster categorizations with the stereotype-inconsistent pairing. The tendency can be found among both whites and African-Americans; but looking at whites alone, the tendency to exhibit faster categorizations with the stereotype-consistent pairing is even more pronounced.[116]

Among other things, the evidence from the IAT provides substantial support for the position advanced by many of the antidiscrimination law commentators noted above that simply creating incentives not to engage in conscious discrimination – as under a conventional prohibition on intentional discrimination – may be of quite limited effect in combating modern employment discrimination.[117] The law is short on effective tools for ferreting out discriminatory behavior that is rooted in unconscious bias, at least when external evidence of discrimination is ambiguous. But while existing reform proposals take the existence of unconscious bias as a given and seek ways to expand the liability attached to such behavior, we are interested in exploring whether the law could help to reduce the bias more directly.

---

[116] See results at https://implicit.harvard.edu/implicit.

[117] See, e.g., Krieger, supra note 30, at 1164-65.

The fact that the bias measured by the IAT is primarily unconscious does not mean that it is not changeable or manipulable. A substantial body of recent social psychology literature strongly questions "the assumption that automatic processes are inflexible and impervious to the perceiver's intentions and goals."[118] Evidence suggests that "both tacit and expressed social influence reduce[] the expression of automatic prejudices" (as measured by the IAT).[119] Might there thus be effective mechanisms by which employment discrimination law could strike directly at the problem of unconscious bias?

Two preliminary comments are important to our discussion. First, our focus is on unconscious bias that produces unlawful discrimination, not on unconscious bias as such. Many people would be troubled by the idea that government should attempt to manipulate people's values simply because it rejects them; we are concerned with the narrower question of how to prevent discrimination in the workplace. We will have more to say on this topic below. Second, while our analysis will point to how both current employment discrimination law and proposed reforms can foster debiasing through law, our claim is not that the law does or could eliminate unconscious bias entirely. Some sources of unconscious bias – such as the very concept of "groupness," which leads individuals to perceive members of their group as more similar to them and members of different groups as more different from them[120] – may not be amenable to debiasing through law. Our claim is only that employment discrimination regimes carry the potential – already realized to some degree – to achieve some important measure of such debiasing.

2. Debiasing Through a Diverse Supervisory Workforce

An intriguing set of results in the social science literature on debiasing of unconscious bias explores the effect of what might be called "role models" or authority figures in the subject's environment on the degree of unconscious bias. One notable study showed that participants who were administered an in-person IAT by an African-American experimenter

---

[118] Irene V. Blair, The Malleability of Automatic Stereotypes and Prejudice, 6 Personality & Soc. Psychol. Rev. 242, 243 (2002).
[119] Brian S. Lowery, Curtis D. Hardin & Stacey Sinclair, Social Influence Effects on Automatic Racial Prejudice, 81 J. Pers. & Soc. Psych. 842, 842 (2001).
[120] See Krieger, supra note 30, at 1186-88.

exhibited substantially less unconscious bias than subjects who were administered an in-person IAT by a white experimenter.[121] In other words, subjects' speed in categorizing black-bad and white-good pairs was closer to their speed in categorizing black-good and white-bad pairs when an African-American experimenter was presiding than when a white experimenter was presiding.

A second study paired white test subjects with African-American experimenters and assigned the pair a task in which (1) the white participant evaluated the African-American experimenter; (2) the white participant needed to cooperate with the African-American experimenter; or (3) the white participant was evaluated by the African-American experimenter. White participants who were told to evaluate the African-American experimenter subsequently exhibited the greatest degree of unconscious bias, while white participants who were evaluated by the African-American experimenter demonstrated the least such unconscious bias (all measured again by speed of categorization of stereotypical versus counter-stereotypical pairs).[122]

These results provide an empirical foundation for the commonsense idea that having a diverse supervisory workforce may well reduce the degree of bias – including unconscious bias – in the workplace. An obvious causal path here is that, compared to supervisory figures who are not members of a particular group, supervisory figures from the group will be less likely to harbor unconscious bias against group members with whom they interact.[123] A recent empirical study by labor economists, for example, finds that African-American hiring officers hire a significantly greater proportion of African-American applicants than do white hiring officers.[124] But the more subtle causal path – and the one we wish to add to the scholarly and policy discussion – is that the simple presence of supervisory figures from the group in question may reduce the degree of unconscious bias exhibited in the workplace by non-group members. And unlike in the laboratory studies

---

[121] Lowery, Hardin & Sinclair, supra note 119.
[122] Jennifer A. Richeson & Nalini Ambady, Effects of Situational Power on Automatic Racial Prejudice, 39 J. Experimental Soc. Psychol. 177 (2003).
[123] Ian Ayres, Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination 419-425 (2001), offers an engaging and insightful discussion.
[124] Michael A. Stoll, Steven Raphael & Harry J. Holzer, Black Job Applicants and the Hiring Officer's Race, 57 Indus. & Lab. Rel. Rev. 267 (2004).

discussed above – where a real worry is that the decline in unconscious bias lasts only for the time of the experiment or a short period longer – in the workplace the presence of a diverse supervisory workforce would be a long-term feature of the individual's environment.

Thus, a logical path for debiasing through employment discrimination law is debiasing through the achievement of a diverse supervisory workforce. How might that be accomplished? The difficulty here is that precisely the existence of the unconscious bias in the first instance suggests the profound barriers standing in the way of achieving a diverse supervisory workforce. The controversial implication, then, is that debiasing through a diverse supervisory workforce may require special *short-term* efforts to achieve diversity among authority figures in the workplace. Title VII's focus on fair treatment of individuals and its disavowal of any mandatory preferences for particular groups – however sensible these twin measures may be from the perspective of solving the problem of conscious employer animus – are simply not apt when the problem is that the very environment and composition of the workplace structures and affects the degree of unconscious bias employment decision makers exhibit. If the degree of unconscious bias is substantial without diverse authority figures and is noticeably lower with them, then a short-term strategy that emphasizes achieving diversity as a central objective may influence the processes that otherwise create the need for an ongoing, long-lived, and invasive program of government regulation of employment relationships, as many believe we currently have under Title VII.

Of course, to some degree government policy in the employment discrimination area already reflects the sort of debiasing through law we are describing here. At all levels of government, public officials have chosen to adopt affirmative action plans governing public sector employees.[125] These initiatives reflect a limited form of the debiasing strategy described here; the plans ensure that there is meaningful diversity across ranks – including supervisory ranks – in the workforce. (Achieving diversity in non-supervisory ranks may be a necessary first step toward creating the pool needed to achieve diversity in supervisory ranks.) Many institutions, private and public, are aware of the value of diverse supervisors as a means of reducing discrimination and improving morale; when the Supreme Court upheld narrowly drawn affirmative action programs in the educational

[125] See, e.g., Johnson v. Santa Clara Transportation Agency, 480 U.S. 616 (1987).

setting, it did so with reference to the arguments of the United States military, stressing points in the same family as those we are offering here.[126] The military's efforts to ensure diversity in its officer corps have been spurred in part by a desire to ensure unity in the ranks and to promote morale among African-American soldiers.[127]

Much broader and more intrusive would be legal mandates by various jurisdictions that private employers, as a matter of employment discrimination law, take definitive steps to achieve diverse supervisory workforces. To be sure, the antidiscrimination laws respond to actual discrimination, not to anticipated discrimination, and hence they cannot be understood to require diverse supervisory ranks as a kind of prophylactic against unconscious discrimination. But in terms of existing law, we can take the courts' willingness to allow voluntary affirmative action with respect to supervisory workers at private firms and to some extent in public institutions as responsive (whether or not intentionally) to the underlying impetus for debiasing of unconscious bias through a diverse supervisory workforce.[128]

3. Debiasing Through Physical and Other Environmental Stimuli

a. Evidence. Social science research on unconscious bias has also suggested how such bias may be reduced by the promotion of counter-stereotypes or elimination of negative stereotypes in the physical or sensory environment. In one study, for instance, participants who spent five minutes creating a mental image of a strong woman showed markedly reduced

---

[126] See Grutter v. Bollinger, 539 U.S. 306, 334 (2003). The brief from the United States military offered a range of claims about the importance of affirmative action in the military; its emphasis was on the importance of diverse supervisors to recruitment and morale. See Consolidated Brief of Julius W. Becton et al., Grutter v. Bollinger (hereinafter "Military Brief"), available at http://www.lexis.com/research/retrieve?_m=fe3b36e48e9f46870631c4f8dff7fd89&docnum=1&_fmtstr=FULL&_startdoc=1&wchp=dGLbVzz-zSkAB&_md5=9fddcec7f0ac5de8f71113be3550099e

[127] See Military Brief, supra note 126, at 16-17.

[128] See United Steelworkers of America v. Weber, 443 U.S. 193 (1979); Grutter 539 U.S. 306.

unconscious bias.[129] In another study, exposure to pictures of counter-stereotypic group members altered unconscious bias; participants who were exposed to photographs of Martin Luther King, Jr. and Timothy McVeigh demonstrated less unconscious bias against African-Americans than subjects exposed to photographs of O.J. Simpson and John F. Kennedy.[130] The results of these studies appear to be a testimonial to the power of the availability heuristic; they also suggest a more general role for the "affect heuristic," by which decisions are formed by reference to a rapid, intuitive, affective judgment about persons, processes, and activities.[131] In fact the affect heuristic undoubtedly plays a role in employment discrimination; and affect can be altered.[132] The general implication is that the *context* in which one individual views another has enormous significance for the nature and degree of unconscious bias.

One fascinating study makes the role of context particularly explicit. In this study, subjects viewed an Asian woman either putting on make-up or using chopsticks (or neither). Those who viewed the woman putting on make-up exhibited substantially more female stereotypes and fewer Asian stereotypes than the control subjects, whereas those who viewed the woman using chopsticks had the opposite response. A similar experiment found that test subjects had less automatic negativity toward African-Americans after viewing an African-American face superimposed onto a picture of the inside of a church than did subjects who viewed the same face superimposed onto a picture of a street. Test subjects had a similar, but muted, reaction to white faces placed in the two contexts.

b. Contexts and biases. These studies about the importance of the surrounding context for the degree of unconscious bias have a resonance with a point familiar to many university students. Students frequently take notice of the portraits of famous scholars or benefactors that adorn classrooms, libraries, offices or other university spaces. In the typical case, the portraits are predominantly white and male. Many students have a strong

---

[129] See Irene V. Blair, Jennifer E. Ma & Alison P. Lenton, Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery, 81 J. Personality & Soc. Psych. 828 (2001).
[130] Dasgupta & Greenwald, supra note 41.
[131] See Paul Slovic, Melissa Finucane, Ellen Peters & Donald G. MacGregor, The Affect Heuristic, in Heuristics and Biases: The Psychology of Intuitive Judgment 397 (Thomas Gilovich et al. eds., 2002).
[132] See id.

experience of these depictions as shaping and reinforcing an environment permeated in subtle ways with various forms of unconscious bias. Recently at Harvard University, for instance, a study revealed that only a small number of portraits at Harvard depict woman, and most of those hang in Radcliffe buildings; this state of affairs led one student leader to comment, "Although it is a minor detail about our campus, it is really a thing that students internalize."[133] "Studying in the Widener Library or eating in Annenberg Dining Hall, Harvard scholars are forever in the company of men – men whose images adorn the University's walls and the halls."[134] Similarly, Harvard students who conducted a comprehensive survey of portraits across campus found that only three of 302 portraits were of persons of color, and reported their findings to the Harvard administration with a request for portraits of "persons of African-American, Asian-American, Latino-American and Native American background, who have served Harvard with distinction." [135]

The social science research just described cannot resolve these complex controversies, but it does suggests the possibility of meaningful effects from, at a minimum, minimizing stimuli such as the Harvard portraits and, more expansively, affirmatively using portraits of important alternative figures, including a larger set of nonwhite and female figures, in their place. Indeed at Harvard, a faculty-student committee was recently charged with choosing portraits of racially diverse figures who have played important roles at Harvard, and these portraits apparently will be placed at significant sites around the university campus.[136]

A similar move is afoot at the United States Capitol. The portrait gallery at the Capitol is "populated almost exclusively by images of white men."[137] In response, Senator Christopher Dodd, the senior Democrat on the Senate Rules Committee, recently engineered the move of one of the few portraits of an African-American, Senator Blanche Kelso Bruce, to a spot "just outside the entrance to the visitors' seats over looking the Senate

---

[133] Kate L. Rakoczy, Men Rule These Walls, Harvard Crimson, June 5, 2003.
[134] Id.
[135] Ken Gewertz, Adding Some Color to Harvard Portraits, Harvard University Gazette, May 1, 2003, at 11.
[136] Id.
[137] Sheryl Gay Stolberg, Face Value at the Capitol: Senator Wants to "Promote Some Diversity" in Congressional Artwork, N.Y. Times, Aug. 13, 2003, at E1.

chamber, in view of the thousands of school children and tourists who pass by each year" – a placement that was "no accident." Likewise, in 1997 a group of lawmakers and advocates for women persuaded Congress to locate a sculpture of three suffragists – Lucretia Mott, Elizabeth Cady Stanton and Susan B. Anthony – in the grand Rotunda at the Capitol.[138] Republican Senator Olympia J. Snowe commented on the placement: "It really talks about the values of our nation and the premium we place on the role of women in our society. Every time I see that statue, I smile, because I think that's where they belong."[139] The effects of such initiatives should not be exaggerated. But behavioral research suggests that Senators Dodd and Snowe are correct to think that reforms of this kind can have real effects on perceptions about particular groups.

Of course – and a similar point applies to the discussion above of debiasing through a diverse supervisory workforce – it is important that the debiasing here not occur in a heavy-handed or aggressive manner. There would presumably be reasonably broad agreement that Mott, Stanton, and Anthony are sufficiently important figures in our nation's history to warrant a central location in the Capitol and indeed that the previous lack of prominent representation of such women may itself not have been a neutral outcome. But a debiasing strategy that replaced representations of important and well-known white, male figures with representations of obscure individuals from other groups would be likely to end up backfiring. (Indeed, the recent effort by Senator Dodd with respect to the Capitol portrait gallery led one conservative critic to remark that the new initiative "let political correctness triumph over accurate history."[140]) An overly aggressive strategy could entrench and increase unconscious bias rather than reducing it. In an analogous move, the influential brief for United States military officers in the *Grutter* litigation[141] took pains to emphasize that in the military, voluntary affirmative action programs ensure that all officers are highly qualified.[142]

In the domain of law, current Title VII law regulates, to a limited extent, the environmental stimuli that surround employees in the workplace. In *Robinson v. Jacksonville Shipyards, Inc.*,[143] for instance, the court held

---

[138] Id. at E5.
[139] Id.
[140] Id. at E1, E5.
[141] Grutter v. Bollinger, 539 U.S. 306 (2003).
[142] See Military Brief, supra note 126, at 29-30.
[143] 760 F. Supp. 1486 (M.D. Fla. 1991).

that pornographic photographs and "pinup" calendars with pictures of nude or partially nude women displayed throughout the work environment could constitute actionable employment discrimination against women.[144] Similarly, in *Waltman v. International Paper Co.*,[145] the court relied on factors including sexually oriented pictures, graffiti, and calendars in the workplace in denying an employer's motion for summary judgment on a sexual harassment claim.[146] And of course, sexual jokes that demean women are a mainstay of Title VII litigation.[147]

The social science literature on debiasing described above provides clear empirical support for these judicial determinations. Just as the depiction of a Chinese woman using chopsticks leads to a heavy focus on her ethnicity while the depiction of the same woman putting on makeup leads to a heavy focus on her sex, the depiction on a poster in the workplace of (for instance) "a prone nude woman with a golf ball on her breast and a man standing over her, golf club in hand, yelling 'Fore!'"[148] may significantly affect employment decision makers' unconscious views of women. Indeed, in a study relied upon by the plaintiffs' expert witness in *Robinson*, men who had viewed a pornographic film just before being interviewed by a woman remembered almost nothing about the interviewer other than her physical characteristics, while men who had watched a regular film before the interview remembered the interview's content.[149] In the words of the *Robinson* court, "The availability of photographs of nude and partially nude women…may encourage a significant proportion of the male population in the workforce to view…women workers as if those women are sex

---

[144] Id.

[145] 875 F.2d 468 (5th Cir. 1989).

[146] Id.; see also Andrews v. City of Philadelphia, 895 F.2d 1469 (3rd Cir. 1990); O'Rourke v. City of Providence, 235 F.3d 713 (1st Cir. 2001); Barbetta v. Chemlawn Services, 669 F. Supp. 569 (W.D.N.Y. 1987).

[147] See, e.g., Robinson v. Jacksonville Shipyards, Inc., 760 F. Supp. 1486, 1504-05 (M.D. Fla. 1991).

[148] Rabidue v. Osceola Refining Co., 805 F.2d 611 (6th Cir. 1986) (Keith, J., dissenting), overruled on other grounds, Harris v. Forklift Sys., 510 U.S. 17 (1993).

[149] Doug McKenzie-Mohr & Mark P. Zanna, Treating Women as Sexual Objects: Look to the (Gender Schematic) Male Who Has Viewed Pornography, 16 Personality & Soc. Psychol. Bull. 296 (1990).

objects."[150] Note that it is certainly possible for those *consciously* biased against women, as well as those unconsciously biased against them, to be affected by physical and environmental stimuli in the workplace. The social science research suggests that removing such demeaning depictions of women is likely to reduce the degree of unconscious bias against women in the workplace.

c. Discrimination, attitudes, and liability. It is important to emphasize that in the existing judicial treatment of sexually explicit material in the workplace, a work environment does not become legally actionable unless its hostility alters women's "terms, conditions, and privileges of employment," and thus constitutes employment discrimination.[151] We do not suggest that the cases find, or should find, actionable employment discrimination solely because demeaning depictions of women may increase unconscious bias in male supervisors or co-workers, apart from the direct effect of these depictions on women's conditions of employment. (An obvious example would be demeaning depictions of women that are consumed privately by one or more male employees out of view of female employees.[152]) No statute authorizes such a strategy of punishing potential thoughts or attitudes apart from a finding of conduct constituting employment discrimination; and any effort to do so would raise obvious First Amendment difficulties.[153]

It is equally clear, however, that prohibiting conduct that is found to alter women's conditions of employment does not become a first amendment violation simply because the prohibition may also accomplish a form of

---

[150] Robinson, 760 F. Supp. at 1503; see also Barbetta, 669 F. Supp. at 573 (stating that the proliferation of sexually explicit material "may be found to create an atmosphere in which women are viewed as men's sexual playthings…").

[151] See Harris v. Forklift Sys., 510 U.S. 17 (1993).

[152] See Jack M. Balkin, Free Speech and Hostile Environments, 99 Colum. L. Rev. 2295, 2316-17 (1999)

[153] In Johnson v. County of Los Angeles Fire Dep't, 865 F. Supp. 1430, 1441 (C.D. Cal. 1994), for instance, the court held unconstitutional on first amendment grounds a fire station policy categorically banning sexually explicit magazines in the workplace; in the court's words, "[The County] may not proscribe the communication of 'sex-role stereotyping' simply because [it] disagree[s] with the message."

debiasing through law.[154] While the first amendment does not ordinarily allow government to regulate speech on the ground that people will be persuaded by it,[155] no plausible interpretation of that amendment forbids the regulation of employment discrimination when the regulation takes the form of forbidding speech that is discriminatory even because of its viewpoint or content.[156] The government may constitutionally impose civil damages against an employer who says, "you're fired because you're a woman" or who repeatedly tells a female employee that she should be posing for Playboy and directs her attention to Playboy posters in the workplace; in either case, the purpose of the law is to prevent an act of discrimination, and the acts in both of these cases can legitimately be defined as such.[157] Like laws that forbid battery and assault, or for that matter purely verbal threats, those that ban discrimination not only have the effect of forbidding certain conduct (including acts that amount to conduct), but also the effect of altering preferences and values. They are not, however, unconstitutional for that reason.

Our central claim is that some acts of discrimination also fuel unconscious bias and hence discrimination. The prospects for debiasing through law provide significant support for some of the existing employment discrimination jurisprudence in this domain. But more controversially, the social science research described above suggests the promise in broader forms of debiasing through employment discrimination law, not limited to the removal of affirmatively demeaning and offensive material as under current Title VII law. We do not mean to recommend imposing affirmative liability under Title VII for maintaining (for instance) portraits that fail to include a diverse set of figures. But an employer's positive effort to portray diversity in the physical environment could be expressly made a factor weighing against

---

[154] For treatments of the constitutionality of existing Title VII doctrine in this area, see Balkin, supra note 152; Richard Fallon, Sexual Harassment, Content-Neutrality, and the First Amendment Dog That Didn't Bark, 1994 Sup. Ct. Rev. 1. As Professor Fallon describes, because the first amendment question was fully briefed in *Harris*, the Court's upholding of hostile environment liability in that case seems to point to a position on the first amendment question. See Fallon, supra, at 1, 5-7.

[155] See generally David Strauss, Persuasion, Autonomy, and the Freedom of Expression, 91 Colum. L. Rev. 335 (1991).

[156] See Balkin, supra note 152, at 2217-18; Fallon, supra note 154, at 21-51.

[157] For an explanation of the conclusion, and on some of the doctrinal complexities here, see Fallon, supra note 154, at 21-51.

employer liability under Title VII – in just the way that, under current Title VII law, employers regularly defend against liability on the basis of actions such as manuals or training videos disseminated in the workplace.[158]

Our basic suggestion is that the existing approach to employer liability might be extended beyond the specific context of the discrete mechanisms contemplated by under present law. If depictions of individuals in the workplace have the significant effects that the empirical research described above suggests they do, then it is only sensible for employment discrimination law – particularly in light of all the complexity it already entails – to attempt to take such effects into account as one factor in determining employer liability under Title VII. Indeed, even without a change in law, it is possible, as Susan Sturm has suggested, that the structures set in motion through existing Title VII doctrines governing employer liability will lead employers to take proactive steps in improving the diversity of physical and other environmental stimuli in the workplace.[159]

C. Debiasing Through Government Information Campaigns

Debiasing through consumer safety law, as explored in section A above, may not be effective or desirable when, for instance, the underlying product is itself illegal in general or for a particular group. In such settings, if the law were able successfully to control the activities of the sellers, typically it would enforce the prohibition on the product, and there would be no need to worry about debiasing the product's consumers. In cases such as this, direct government intervention in the form of legislated or administratively-driven information campaigns may be a desirable means of debiasing through law. The general concept of debiasing through government information campaigns is of course familiar,[160] but we mean to illustrate it here with an important example of such debiasing that has not been discussed in the legal literature. We also use the analysis of debiasing through government information campaigns to illustrate several important points in our normative analysis in Part III below.

---

[158] See Faragher v. City of Boca Raton, 524 U.S. 775 (1998); Kolstad v. American Dental Assn., 527 U.S. 526 (1999).

[159] See generally Sturm, supra note 110.

[160] See, e.g. Jeffrey J. Rachlinski, The Uncertain Psychological Case for Paternalism, 97 Nw. U. L. Rev. 1165, 1224 (2003).

In settings in which the government wishes to discourage use of a product, debiasing through law strategies may profitably harness what has recently been termed in the public health field the "social norms" approach, which draws upon the large role played by norms in governing decisions about whether to engage in illegal activities.[161] Consider, for instance, the problem of alcohol abuse by (mostly-underage) college students. According to a recent survey by the Harvard School of Public Health, about forty-four percent of college students engaged in binge drinking in the two-week period preceding the survey.[162] (Binge drinking is defined as five drinks or more in a row for men and four drinks or more in a row for women.[163] Note again that alcohol consumption is unlawful for most of these students.) Why does binge drinking occur? There is no single answer, but evidence suggests that a particular bias – student overestimation of the level of drinking on campus – plays a substantial role.[164] Most students believe that alcohol abuse is far more pervasive than it actually is. Misperceptions of this kind result, in large part, from the availability heuristic.[165] Incidents of alcohol abuse are easily recalled, and the consequence is to inflate perceptions. The bias in estimation has significant effects. College students are generally affected by their beliefs about what other college students do, and hence alcohol abuse will inevitably increase if students have an inflated perception of reality with regard to other students' drinking behavior.[166]

In such circumstances, it is possible to believe that a strategy of debiasing through law, counteracting students' reliance on the availability heuristic, will actually lead to significant changes. If government seeks to reduce alcohol abuse by college students, it might do well to emphasize the statistical reality. (For obvious reasons, private producers of alcoholic beverages would not be in a good position to publicize information about the occurrence of underage drinking, and government probably would not want to encourage this.) Alert to the possibility of achieving debiasing through government informational campaigns emphasizing the statistical reality, many government actors have adopted a specific approach – the "social norms" approach – for correcting student beliefs about the pervasiveness of

---

[161] See generally H. Wesley Perkins, The Social Norms Approach to Preventing School and College Age Substance Abuse 7-8 (2003).
[162] http://www.hsph.harvard.edu/cas/rpt2000/CAS2000rpt.shtml.
[163] Id.
[164] See Perkins, supra note 161, at 8-9.
[165] Id. (telling an availability story, although not explicitly using the term).
[166] Id. at 8-9.

the relevant behavior. Montana, for example, has adopted a large-scale social norms campaign, one that has stressed the fact that strong majorities of citizens of Montana do not drink.[167] One advertisement attempts to correct misperceived norms on college campuses by asserting, "Most (81 percent) of Montana college students have four, fewer, or no alcoholic drinks each week."[168]

Montana applies the same approach to cigarette smoking (where again it is easy to see why producers of the item in question would not be in a good position to publicize information about teen smoking, and why the government would not want to encourage this). Montana's smoking advertisement asserts that "Most (70 percent) of Montana teens are tobacco free." [169] In this context there is data showing that the strategy has produced statistically significant improvements in the accuracy of social perceptions and also statistically significant decreases in smoking.[170]

It is true that much of the evidence of reduced substance abuse from these studies of debiasing through government information campaigns comes from individuals' self-reports; the difficulty with such self-reports is that they may move in response to the government information campaign even if underlying behavior is unchanged. One of the studies in this area, however, examines not self-reports of social perceptions and substance abuse but actual arrests for liquor law violations. The study found a forty-six percent drop in arrests, from eighty-four to forty-four per year, after the government information campaign was instituted.[171] Thus, there is reason to think that the

---

[167] See Jeffrey W. Linkenbach, The Montana Model: Development and Overview of A Seven-Step Process for Implementing Macro-Level Social Norms Campaigns, in H. Wesley Perkins, The Social Norms Approach to Preventing School and College Age Substance Abuse 182 (2003).
[168] Id. at 195.
[169] Jeffrey Linkenbach & H. Wesley Perkins, MOST of Us Are Tobacco Free: An Eight-Month Social Norms Campaign Reducing Youth Initiation of Smoking in Montana, in H. Wesley Perkins, The Social Norms Approach to Preventing School and College Age Substance Abuse 224, 230-33 (2003).
[170] See id.
[171] H. Wesley Perkins & David W. Craig, The Hobart and William Smith Colleges Experiments: A Synergistic Social Norms Approach Using Print, Electronic Media, and Curriculum Infusion to Reduce Collegiate Problem Drinking, in H. Wesley Perkins, The Social Norms Approach to Preventing School and College Age Substance Abuse 35, 61 (2003).

"social norms" approach – using statistical reality to debias a target audience – produces actual effects on usage of illegal products.

The apparent success of such debiasing strategies in the substance abuse context is particularly notable given the lack of results associated with other, more familiar strategies to decrease the frequency of such behavior. Educational efforts emphasizing the health risks, for instance, have tended to produce small or no effects, as young people often dismiss their own odds of facing serious harm.[172] And attempts to encourage attendance at "alternative social events" on campus have (not too surprisingly) been similarly ineffective.

Beyond the substance abuse context, debiasing through government provision of information about others' behavior may have considerable promise for reducing the frequency of other actions that are harmful to self or others. Suppose, for instance, that the availability heuristic is a contributing factor in leading people to violate the tax laws (because media reports of cheaters raise estimates of the frequency of cheating) or fall victim to eating disorders,[173] on just the same theory as was described above in connection with substance abuse. If so, government information campaigns might be able to correct the error and to alter behavior. In the context of tax compliance, a real-world experiment conducted by officials in Minnesota produced exactly this effect.[174] Apparently some taxpayers are more likely to violate the law because of a misperception – plausibly based on the availability of media or other accounts of cheaters – about the level of noncompliance. When informed that the actual compliance level is high, they are debiased, and are less likely to cheat.[175] By contrast, when they are told that "your income tax dollars are spent on services that we Minnesotans depend on. Over 30 percent of state taxes go[es] to support education.

---

[172] Cf. Armour & Taylor, supra note 67, at 334.

[173] See Jeffrey W. Linkenbach, H. Wesley Perkins & William DeJong, Parents' Perceptions of Parenting Norms: Using the Social Norms Approach to Reinforce Effective Parenting, in H. Wesley Perkins, The Social Norms Approach to Preventing School and College Age Substance Abuse 247, 266-69 (2003).

[174] Stephen Coleman, The Minnesota Income Tax Compliance Experiment State Tax Results 5-6, 18-19 (1996), available at http://www.state.mn.us/ebranch/mdor/reports/compliance/pdf.

[175] See id.

Another 18 percent is spent on health care and support for the elderly and needy," their compliance levels are unaffected.[176]

D. Debiasing Through Liability Rules

As noted above, the social science literature on debiasing of boundedly rational actors has focused heavily on steps that reduce the effects of judgment biases as distinguished from departures from expected utility theory. Indeed, social scientists have paid little attention to the debiasing of departures from expected utility theory. The reason may be that such departures are not unambiguous "errors," and thus it is controversial to say (for example) that the endowment effect, or loss aversion, is a kind of mistake that requires correction.

A substantial literature in law addresses this issue, exploring the central empirical finding that there is a substantial difference between individuals' willingness to accept – the amount at which someone would sell an entitlement – and their – willingness to pay – the amount this individual would pay to purchase the entitlement.[177] Contrary to the prediction of the Coase theorem and consistent with the endowment effect, the empirical evidence shows that people require far more to give up a good that they hold than they are willing to pay to obtain the good in the first instance.[178] For purposes of law, the critical question is whether willingness to accept or willingness to pay should be taken the preferred measure of value. Among other things, the answer to this question bears on damages in tort law[179] and on the appropriate measure of goods for purposes of environmental policy and cost-benefit analysis.[180]

---

[176] Id. at 5.

[177] See, e.g., Russell Korobkin, Note, Policymaking and the Offer/Asking Price Gap: Toward a Theory of Efficient Entitlement Allocation, 46 Stan. L. Rev. 663 (1994) (surveying empirical research and offering further analysis of the normative question).

[178] See Kahneman, Knetsch & Thaler, supra note 33, at 1329-42.

[179] See Edward J. McCaffery, Daniel J. Kahneman & Matthew L. Spitzer, Framing the Jury: Cognitive Perspectives on Pain and Suffering Awards, 81 Va. L. Rev. 1341, 1353-54, 1364-73 (1995).

[180] See generally Jack L. Knetsch, Reference States, Fairness, and Choice of Measure to Value Environmental Changes, in Environment, Ethics, and Behavior 13 (Max H. Bazerman et al. eds., 1997).

Our analysis in this section stipulates that there are contexts in which, when willingness to accept and willingness to pay differ, the latter is the normatively preferred measure of value. Imagine, for instance, a situation in which the endowment effect leads individuals to be excessively attached to an entitlement that ideally should trade fairly freely. The excessive attachment may be a product of bargaining considerations; perhaps people are not willing to trade certain goods because they are acting strategically.[181] Alternatively, the excessive attachment may stem from a failure to appreciate the opportunity costs of refusing to trade the good in question[182]; this failure can plausibly be seen as a form of bounded rationality, leading people to be reluctant to trade in circumstances in which they ought to do so. It is also possible that a high willingness to accept reflects a judgment that it is morally questionable to sell certain goods (such as environmental amenities) for "any" price[183]; let us simply assume that contexts can be found in which this judgment is hard to defend, if only because the relevant amounts can be used to provide or to protect other environmental amenities. In such situations, eliminating the endowment effect through a lowering of individuals' willingness to accept to the level of their willingness to pay is a desirable step.[184] Might a strategy of debiasing through law be able to accomplish this?

An empirical study by Jeffrey Rachlinski and Forest Jourden suggests that the divergence of willingness to accept from willingness to pay is importantly affected by the way in which an entitlement is protected from violation.[185] In Guido Calabresi and Douglas Melamed's classic treatment, entitlements can be protected either by liability rules, which allow forced acquisition of rights in exchange for payment of damages, or property rules,

---

[181] See Russell Korobkin, The Endowment Effect and Legal Analysis, 97 Nw. U. L. Rev. 1227, 1243-44 (2003).

[182] See Cass R. Sunstein, Switching the Default Rule, 77 N.Y.U. L. Rev. 106, 131 (2002).

[183] See Kahneman, Knetsch & Thaler, supra note 33, at 1327 tbl.1.

[184] We do not intend to resolve here the question whether willingness to pay is always or generally a preferable measure to willingness to accept. Undoubtedly a high willingness to accept is sometimes a product of a strong attachment to the good in question, perhaps as a result of learning.

[185] Jeffrey J. Rachlinski & Forest Jourden, Remedies and the Psychology of Ownership, 51 Vand. L. Rev. 1541 (1998).

which prevent compelled alienation through injunctive enforcement.[186] Of course the law – especially the law of property – is generally in the business of choosing whether to use liability rules or property rules. Does the choice affect valuation?

Rachlinski and Jourden's study finds a marked reduction in the endowment effect, and hence the disparity between willingness to accept and willingness to pay, in circumstances in which liability rules rather than property rules protect the entitlement in question.[187] In the standard endowment effect pattern, willingness to accept is well above willingness to pay in the property rule situation. But in the liability rule situation, both willingness to accept and willingness to pay are at the same level at which willingness to pay was in the property rule situation. Rachlinski and Jourden offer an explanation of their results by suggesting that "a right that is protected by a damages remedy might convey less of a sense of ownership than does a right that is protected by an injunctive remedy."[188] Such incomplete ownership prevents a perfection of the emotional attachment that is harbinger of the endowment effect.

As Ian Ayres suggests in an article following Rachlinski and Jourden's work, their empirical findings suggest that in areas of trade in which lawmakers want to facilitate transactions, liability rules may be preferable to property rules.[189] Applying our framework, choosing liability rules over property rules can be regarded as a form of debiasing through law. Liability rules, under Rachlinski and Jourden's findings, eliminate the endowment effect by moving individuals' willingness to accept down to the level of their willingness to pay.[190]

---

[186] Guido Calabresi & A. Douglas Melamed, Property Rules, Liability Rules, and Inalienability: One View of the Cathedral, 85 Harv. L. Rev. 1089 (1972).
[187] Rachlinski & Jourden, supra note 185, at 1566.
[188] Id. at 1560.
[189] Ian Ayres, Protecting Property with Puts, 32 Val. U. L. Rev. 793, 811-12 (1998).
[190] These remarks about the possibility of eliminating the endowment effect are in line with the important finding that while people show that effect when acting in their individual capacities, they do not do so when acting in the role of corporate manager in a business agency context. See Jennifer Arlen, Matthew Spitzer & Eric Talley, Endowment Effects Within Corporate Agency Relationships, 31 J. Legal Stud. 1 (2002).

It remains to be seen whether, across a range of contexts, the device of choosing liability rules over property rules brings willingness to accept into line with willingness to pay. (Recall that we are focusing on situations in which the latter measure is normatively preferred if the two measures conflict.) Of course people are often unaware of how, exactly, their entitlements are protected; if the legal system uses liability rules rather than property rules, most people will not be aware of it. Note also that in Rachlinski and Jourden's study, the entitlements involved environmental amenities. In that distinctive context, the occurrence of the endowment effect under a property rule may have been "motivated by subjects' belief that it is improper to sell an environmental resource that one can protect" while this belief was not triggered under a liability rule "because the law permitted the destruction of the resource for a price."[191] Absent the societal commitment to environmental amenities, for which people often demand a great deal (and on occasion refuse to sell at any price at all),[192] it remains possible that the choice between property and liability rules would not have the same impact on willingness to accept versus willingness to pay.[193]

E. Debiasing Through Changes in Default Rules

Sometimes individuals show biased judgments not about risky products, members of other groups, or their peers (as we have emphasized in the preceding sections focusing on judgment biases) but rather about the content of the law itself. Like unconscious bias on the part of employers, such biased legal judgments have not been considered in the heuristics-and-biases literature. But they similarly lead to systematic errors of the sort potentially conducive – because of their predictability – to effective responses based on behaviorally informed analysis. In the present section we ask how debiasing through law might be used to correct erroneous understandings of the law's content.

---

[191] Korobkin, supra note 181, at 1285.

[192] See Kahneman, Knetsch & Thaler, supra note 33, at 1327 tbl.1.

[193] Daphna Lewinsohn-Zamir, The Choice Between Property Rules and Liability Rules Revisited: Critical Observations from Behavioral Studies, 80 Tex. L. Rev. 219, 250-57 (2001), argues, but without presenting any direct empirical evidence, that property rules may be preferable to liability rules for reducing the endowment effect in some contexts.

Consider, for purposes of the discussion here, the question of discharge from employment. In this context workers appear to have dramatically mistaken beliefs about whether firms are legally permitted to terminate their employment absent good cause. Evidence shows that workers systematically overestimate the degree of protection the legal system affords them.[194] Pauline Kim has found that most workers falsely believe that employment can be terminated only for good cause.[195] In Missouri, for instance, extremely strong majorities of employees – eighty percent or more – believed that the following grounds for discharge, entirely lawful in Missouri, are in fact unlawful: the employer wants to hire someone else to do the same job; the employer mistakenly believes that the employee stole money; or the employer personally dislikes the employee.[196] Similar results were found in California and New York, notwithstanding substantial variations in the law of the three states.[197] Kim's most general result is that overwhelming majorities of workers falsely believe that discharges that fall short of "good cause" (based on a job-related reason) are prohibited by law. This ignorance on the part of workers cuts across distinctions that might be thought to make a difference – not only geography, but also age, work experience, and union experience.

These worker misperceptions might naturally be seen as a simple absence of information. But because the errors consistently go in a single direction – an exaggerated sense, by workers, of their legal rights – we think it is useful to conceptualize the behavior in question as a form of bias. People often reduce cognitive dissonance by drawing their beliefs about how things should be in line with their views about how things are.[198] This phenomenon suggests a role for what might be termed the "legal fairness heuristic," in accordance with which judgments about legal rights are influenced by judgments about what the law should be. If workers' beliefs about what the law is tend to reflect their beliefs about what the law ought to be, then they

---

[194] Richard B. Freeman & Joel Rogers, What Workers Want 118-21 (1999): Pauline Kim, Bargaining With Imperfect Information: A Study of Workers Perceptions of Legal Protection in an At-Will World, 83 Cornell L. Rev. 105, 133-46 (1997).
[195] Kim, supra note 194, at 134.
[196] Id.
[197] Pauline Kim, Norms, Learning, and Law, 1999 U. Ill. L. Rev. 447, 451.
[198] See generally George Akerlof & William T. Dickens, The Economic Consequences of Cognitive Dissonance, in George Akerlof, An Economic Theorists' Book of Tales 123 (1984).

will sometimes systematically misperceive the law, in a way that reflects their normative judgments. The mechanisms here are not entirely clear. For our purposes, what is most important is that workers' beliefs with respect to their legal rights will show large-scale errors in a predictable direction if a legal rule is in conflict with widespread moral intuitions. Again, there is important overlap here with problems of imperfect information, but that category – which embraces, for instance, situations in which people have no opinion or prediction about matters on which they lack good information – sweeps much more broadly than the specific phenomenon we are seeking to describe here.

A modest response to the current state of affairs regarding workers' perception of valid grounds for discharge would involve debiasing through law in the form of a simple shift in the default rule governing discharge. Samuel Issacharoff suggested such a reform some years ago.[199] This shift would be specifically designed to counteract employees' misperceptions of their legal rights. If the legal rule is changed so that employees have protection against discharge without good cause unless they expressly agree otherwise, then workers are more likely to learn the actual state of affairs, whether or not the parties decide to contract around the default rule. (Again, the overlap with informational analysis is clear.) In fact a mild but unmistakable movement in this direction, and hence a form of implicit debiasing through law, can be found in the judicial decisions taking ambiguous employer statements as creating legal protection against discharge without good cause.[200]

Thus, for instance, the Wyoming Supreme Court was faced with a case in which an employee handbook both described detailed procedures the employer would follow in the event of employee performance problems and stated on the first page that the handbook "was not an employment contract." [201] The employee who was threatened with termination without the specified procedures having been followed had signed an application form containing

---

[199] Samuel Issacharoff, Contracting For Employment, 74 Tex. L. Rev. 1783 (1996).

[200] For a general discussion, see Comment, Judicial Interpretation of Employee Handbooks: The Creation of a Common Law Information Eliciting Penalty Default Rule, University of Chicago Law Review (forthcoming 2004).

[201] McDonald v. Mobil Coal Producing, Inc., 789 P.2d 866 (Wyo. 1990), reversed, 820 P.2d 986 (Wyo. 1991).

the following disclaimer: "READ CAREFULLY BEFORE SIGNING. I agree that any offer of employment, and acceptance thereof, does not constitute a binding contract of any length, and that such employment is terminable at the will of either party, subject to appropriate state and/or federal laws."[202] The court held that the attempt to disclaim legal limits on termination was ineffective notwithstanding this express statement. This holding, otherwise surprising, makes behavioral sense. In light of Kim's findings, most employees would tend to take the ambiguous employer provisions to fortify their belief that they have a right to be discharged only for good cause, the opposite of the actual regime of at-will employment.

In a separate line of cases, also responding to likely worker assumptions, courts sometimes take oral promises to be binding, even if they have a degree of ambiguity. In *Toussaint v. Blue Cross*,[203] for example, the court was confronted with statements from personnel officers that an employee would be able to continue to work "as long as I did my job," and or as long as "I was doing the job."[204] In an alternative holding, the court held that these statements were sufficient to justify an action for wrongful discharge, on the theory that the ambiguous oral statements contractually modified the usual at-will regime.

The Model Employment Termination Act (META)[205] takes the basic idea much further. It creates a right to protection from discharge in the absence of good cause, but allows employers and employees to waive the right, on the basis of an agreement by the employer to provide a severance agreement in the event of a discharge not based on good cause.[206] Montana, alone among the fifty states, has adopted a statute that, like META, requires good cause for discharge – but Montana's requirement is essentially not waivable, and hence that state goes well beyond a change in the default rule.[207]

Debiasing through changing the default rule, as discussed in this section, differs in significant respects from the other examples of debiasing

---

[202] McDonald v. Mobil Coal Producing, Inc., 820 P. 2d 986, 987 (Wyo.1991).
[203] 408 Mich. 579 (1980).
[204] Id. at 610.
[205] Model Employment Termination Act (1991).
[206] Id. § 4(c).
[207] See Montana Wrongful Discharge from Employment Act, Mont. Code Ann. §§ 39-2-901 to 914 (2002).

through law discussed above. In our previous illustrations, as well as in discussions of debiasing through law (in the form of procedural rules) in the existing legal literature, the law attempts to counteract people's biases by altering those judgments to bring them in line with reality. A parallel approach, in the case of biased judgments about the law, would be to take steps to alter those biased judgments. Our suggestion here is different. Given the empirical evidence of persistent unawareness of the at-will rule, debiasing through law in this context takes the alternative form of bringing legal reality into line with employees' judgments. But the difference should not be overstated. Because only default rules are involved, the effect of the shift may be to change workers' perceptions – just as in the usual type of debiasing through law – simply because employers will often respond by unambiguously stating that employment is at-will.

In Michigan, for instance, a major effect of decisions that move toward a change in the default rule may have been to encourage employers to tell employees, in the plainest terms, that they may be discharged for any reason or for no reason at all.[208] In fact we can make good behavioral sense of some courts' insistence that such disclaimers be exceedingly clear. Kim demonstrates that many employees are likely to misunderstand most disclaimers or to believe them to be ineffective.[209] An extremely clear statement, made necessary by judicial decisions, can be understood as a way of overcoming this bias on the part of employees. It is even possible – hearkening back to our analysis in section A – that an extremely clear but highly general statement is not enough to achieve debiasing of individuals with a robust "legal fairness heuristic." In that case requiring a direct and particular statement, along the lines discussed in section A in the consumer safety context, might be warranted in lieu of a simple change in default rule.

We do not contend that a finding of worker misperceptions is necessarily sufficient to justify a change in the default rule governing discharge from employment, or even to support all of the judicial decisions described above. Perhaps arbitrary discharges are rare; and if so, then there might be no large problem for the legal system to solve. To evaluate any claim for debiasing through law, it is necessary to know whether a demonstrated bias has important effects. Conversely, it is possible that a change in the default rule is too tepid a response to the at-will regime. Once again, however, our debiasing through law approach charts a middle ground

---

[208] See, e.g., Reid v. Sears, Roebuck & Co., 790 F.2d 453 (6th Cir. 1986).
[209] Kim, supra note 194, at 137.

between the less and more aggressive regulatory strategies. And the approach is not limited to employees. Suppose that consumers also have an exaggerated understanding of their legal rights; if so default rules, and principles governing waiver, might be used as a way to protect consumers by debiasing their perceptions of law as well.

F. Debiasing Through Corporate Law

Our opening discussion referred to an example of "debiasing law" in the corporate area – imposition of the business judgment rule as a means of guarding against hindsight-biased decisions by adjudicators considering liability of corporate actors.[210] But corporate law also provides a recent intriguing example of debiasing through law, in which substantive law is structured not to work around a specified form of boundedly rational behavior (which itself is taken as a given) but, instead, to work to reduce or even eliminate the boundedly rational behavior.

Much recent scholarship in the corporate law area has been concerned with the question of the optimal breakdown of board composition between so-called "inside" and "outside" directors.[211] Inside directors are those who are primarily employed by or otherwise closely connected with the corporation; outside directors, by contrast, have no such close link to the firm. A number of arguments support the inclusion of at least some outside directors on the board.[212] Of particular relevance for our purposes is Donald Langevoort's suggestion – although he ultimately does not join those pressing for further increases in outside directors – that the involvement of such directors may help to overcome optimistically biased judgments ("organizational optimism") on the part of inside directors.[213]Langevoort's suggestion is strengthened by recent arguments, in the wake of the Enron

---

[210] See Rachlinski, supra note 4, at 620-21; supra notes 4-5 and accompanying text.

[211] For a recent summary, see Donald C. Langevoort, The Human Nature of Corporate Boards: Law, Norms, and the Unintended Consequences of Independence and Accountability, 89 Geo. L.J. 797, 797-99 (2001).

[212] See Ronald J. Gilson & Reiner Kraakman, Reinventing the Outside Director: An Agenda For Institutional Investors, 43 Stan. L. Rev. 863, 873 (1991); Robert C. Pozen, Institutional Investors: The Reluctant Activists, Harv. Bus. Rev., Jan.-Feb. 1994, at 140.

[213] See Langevoort, supra note 211, at 803, 809.

debacle, for more disagreement and dissent on corporate boards.[214] Those arguments make a great deal of sense in light of the finding that the probability of erroneous decisions is increased when deliberations are undertaken by like-minded people.[215] Often those who agree with one another will end up at a more extreme point in line with their predeliberation tendencies. In the context of corporate boards, the prediction is that optimistic members will lead one another in the direction of further optimism and excessive risk-taking. As a result, boards might well end up more optimistic than the median board member before deliberation began; if so, boards will almost inevitably blunder.

Of course, it is possible that market pressures will impose meaningful constraints on the degree of optimism bias exhibited by inside directors or corporate boards.[216] But at the same time, other forces may tend to increase the degree of optimism bias such individuals exhibit. These include the process by which such executives are selected and the link between these individuals' optimistic judgments and their self-conception and esteem.[217]

Why might requiring the involvement of outside directors on corporate boards help to reduce the degree of optimism bias exhibited by inside directors? One reason is that the selection of outside directors is not necessarily heavily influenced by whether candidates have highly optimistic views of the firms' prospects (in contrast to the case of top executives' selection[218]). Another is that outside directors' self-conception and esteem are, relative to the case of inside directors, less closely bundled up with the firm's fortunes.[219] Outsider directors might well serve to check deliberative

---

[214] See Jeffrey A. Sonnenfeld, What Makes Great Boards Great, Harvard Business Review (Sept. 2002).
[215] See generally Cass R. Sunstein, Why Societies Need Dissent (2003).
[216] Avishalom Tor, The Fable of Entry: Bounded Rationality, Market Discipline and Legal Policy, 101 Mich. L. Rev. 482 (2002), provides analysis of the effects and limits of market pressures as a constraining force in the context of firm entry into new industry.
[217] Langevoort, supra note 211, at 809; Donald C. Langevoort, Organized Illusions: A Behavioral Theory of Why Corporations Mislead Stock Market Investors (And Cause Other Social Harms), 146 U. Pa. L. Rev. 101, 140 (1997).
[218] See Langevoort, supra note 211, at 809.
[219] Id. at 803.

processes that fuel unrealistically optimistic decisions; even a single dissenter is often able to move such processes in better directions.[220]

Consider in this light the expanded requirement under the Sarbanes-Oxley Act that boards use outside directors to perform all auditing functions (so that a threshold number of outside directors must be named to the board).[221] We might well see this requirement as a means of debiasing through law by responding directly to the risk of unrealistic optimism on boards. Corporate law governing the structure of the legal-organizational form of the board of directors may be a reasonable way to reduce the degree of bias exhibited by inside directors. Of course, it is also possible that in some cases companies would exercise self-help to achieve the same effect (putting outside directors on the board without a legal requirement). In the corporate context, many boards do contain some outside directors, and this may reflect in part a self-help step by firms interested in combating problems of optimism bias. A legal requirement such as Sarbanes-Oxley, however, is likely to facilitate such debiasing on a broader scale, although at a cost of requiring outside directors on all covered boards notwithstanding substantial firm- and industry- specific variation in ideal board structure.[222]

III. Normative Issues

Our analysis in Part II of debiasing through substantive law suggests an important complement to the various forms of debiasing through procedural rules discussed in the existing legal literature. "[G]overnments can adopt measures that restructure decisions as a less intrusive alternative to paternalistic restrictions on choice."[223] Debiasing through substantive law may provide a more direct and effective response to problems of bounded rationality than the more typical approach of "debiasing law." But whether debiasing occurs through procedural rules or through substantive law, important normative issues may arise from the use of debiasing strategies. The central reason is that in using such strategies, the government is often

---

[220] See Sunstein, supra note 215.
[221] Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, § 301.
[222] On the importance of such variation, see Langevoort, supra note 211, at 815.
[223] Rachlinski, supra note 160, at 1224. Rachlinski offers the general observation but does not develop it with reference to prospects for debiasing through substantive law.

deliberately and self-consciously engaged in altering people's values and their perceptions of the world around them. A threshold question raised by many versions of debiasing through law is whether and when the government is appropriately involved in this task.

A. Debiasing Everywhere?

A starting observation in answering this question is that some government actions that influence individual values and perceptions are entirely uncontroversial. Criminal and civil law, designed to forbid force and fraud, are the most elementary forms of "debiasing"; they have the effect of discouraging people from believing that it is appropriate to engage in force and fraud. In fact that form of "debiasing" is a large part of the basic point of prohibitions on force and fraud, and hence it would be odd to suggest that government must, with respect to these forms of wrongdoing, remain neutral about people's values. The same can (and has) been said for the basic institutions of a market economy, including freedom of contract and private property. As Albert Hirschman has demonstrated, freedom of contract and its accompanying institutions have important value-shaping purposes and effects. If people see one another as trading partners, they are more likely to cooperate with one another, as captured in the notion of "doux commerce."[224] Hirschman's account of capitalist institutions is essentially one of debiasing, as socially destructive passions, involving differences of religion and ethnicity, become less important and less damaging under conditions in which people follow their material interests instead.

A clear modern analogue is employment discrimination law, discussed in Part II.B. Insofar as such law forbids racist or sexist behavior, it is intended to reduce the extent and effects of certain preferences and beliefs. Debiasing, in its literal form, lies at the heart of prohibitions on discrimination on the basis of race, sex, and other traits. Of course government targets actions, not what is inside heads. It would be possible to insist that the relevant prohibitions are directed against behavior rather than values or beliefs; the law does not direct itself against values or beliefs as such. The point is correct. But prohibitions on action affect, and are often intended to affect, values and beliefs as well. If we neglect the value-shaping character of employment discrimination law, we will misconceive its purpose and effect. When government announces that it will punish employers for

[224] See generally Albert Hirschman, The Passions and the Interests (1977).

their employees' discriminatory behavior, it may encourage, as Susan Sturm has argued, various steps including those that both subtly and significantly shape employees' attitudes,[225] but no one suggests that employment discrimination law therefore is inappropriately manipulative or otherwise normatively objectionable. As we described above, both current employment discrimination doctrine (governing creating of hostile work environments) and our suggested elaboration of current doctrine (governing the determinants of employer liability under Title VII) are illustrations of debiasing through law but do not appear to raise any new or special normative issues.

B. Factual Errors and Paternalism

We have emphasized that it is objectionable, and usually unconstitutional, if government bans speech on the ground that people are likely to be persuaded by it; that particular form of debiasing through law is constitutionally unacceptable.[226] But others forms of debiasing are perfectly legitimate, even if they involve efforts to alter people's perceptions of reality. In countless domains the government either discloses information on its own or requires disclosure by those providing goods or services.[227] In cases in which people are committing a clear factual error, and in which the government's strategy is a straightforward effort to meet falsehood with fact, these interventions command substantial agreement. It is hard to think of a plausible normative objection to strategies of this kind.

Many instances of debiasing through substantive law fit just this unobjectionable pattern. Where, for instance, the availability heuristic leads individuals to exaggerate the number of peers who engage in underage drinking, debiasing through substantive law takes the form of emphasizing the statistical reality with the goal of correcting people's misimpressions, as discussed in Part II.C above. While the shortcoming that gives rise to the government intervention is, in the account of debiasing through law, a bias rather than a mere absence of information, the government response is identical. As in the context of information disclosure generally, the normative

---

[225] See generally Sturm, supra note 110.

[226] See supra note 155 and accompanying text.

[227] See, e.g., W. Kip Viscusi & Wesley A. Magat with Joel Huber et al., Learning About Risk: Consumer and Worker Response to Hazard Information (1987); Bradley Karkkainen, Information As Environmental Regulation, 89 Geo. L.J. 257 (2001).

question here does not require investigation of controversial questions about individual autonomy and the risk of manipulation. The real question is whether the effort at correction is effective and, if so, whether it is cost-justified. The same analysis applies to our treatment in Part II.D above of switching the default rule governing discharge from employment in response to most individuals' misperceptions of their rights to maintain their jobs in the absence of good cause for termination.

Of course some people, intuitively or reflectively alert to the risk of bias, correct their own errors.[228] Recall our earlier discussion of two cognitive systems, the heuristic-driven System I and the more deliberative and calculative System II.[229] As we suggested, people often use their own System II to correct the operation of System I.[230] In some cases – for instance through "cooling off periods" in consumer contexts – the law may even have an important role to play in facilitating such reliance on System II.[231] We are suggesting not that deliberative self-correction is rare or impossible, but that systematic errors are widespread, and that many government efforts at debiasing through law in response to such errors do not run afoul of prohibitions on official manipulation.

An important feature of debiasing through law, quelling normative doubts, is that the primary effects of the action are likely to be felt by those whose errors prompted government involvement in the first place. When government action takes the form of providing accurate information, it is reasonable to expect that the action will not introduce significant new distortions in the behavior of those who were not subject to either information failures or judgment biases in the first instance.[232] In this respect, provision of accurate information in response to either a conventional information failure or a bias fits well with a broader emerging theme in prescriptive work in behavioral law and economics: *Adopt approaches that will correct errors, but without imposing significant costs on those who are unlikely to err*. Colin Camerer, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin, for instance, have argued on behalf of an

---

[228] See, e.g., Rachlinski, supra note 160, at 1211-19.
[229] See supra notes 46-47 and accompanying text.
[230] See Kahneman & Frederick, supra note 22, at 51.
[231] See Camerer, Issacharoff, Loewenstein, O'Donoghue & Rabin, supra note 14, at 1240-42; Rachlinski, supra note 160, at 1224.
[232] See Camerer, Issacharoff, Loewenstein, O'Donoghue & Rabin, supra note 14, at 1232-35.

"asymmetrical paternalism," that is, a strategy that counteracts errors that reduce welfare, but that does not significantly affect people who did not previously err.[233] Existing law in the consumer credit context, for example, reflects such an approach in its requirement that lenders disclose particular facts such as the total interest payment over the life of a loan.[234] This sort of approach, applied to the context of judgment biases, recognizes the important truth that not all individuals are likely to be boundedly rational, at least to the same degree.[235] Debiasing through law in these settings is a form of asymmetrical paternalism in the sense urged by Camerer and his coauthors.

Even beyond cases in which debiasing through substantive law is simply an effort to meet falsehood with fact, such debiasing will frequently satisfy the goal of mostly affecting those likely to err. Many of the strategies for debiasing of boundedly rational actors discussed above have this character. The limited Title VII step of regulating sexually explicit material displayed in the workplace, for example, should not seriously affect those who do not suffer from unconscious sexism in the first instance (although this is not to say that they could never be burdened by this application of Title VII). Of course, in all of our examples, the government intervention may not be entirely cost-free for those who did not previously exhibit bounded rationality. Indeed, the mere provision of statistically accurate information by government – either to correct a simple absence of information or to respond to judgment bias – may impose costs on those who did not err prior to the intervention simply because of the burden of processing the information. But if the government intervention produces important benefits for those who are prone to decision making errors, then on balance the intervention may be desirable.

The absence of significant effects on those whose behavior is not in need of correction marks a substantial contrast between some of the strategies for debiasing through law discussed above and their counterpart "debiasing law" approaches. Suppose, for example, that one responds to the hindsight bias not by trying to reduce the occurrence of biased decision making but by facilitating insulation of legal outcomes from the effects of such decision making through the use of the highly deferential business judgment rule.[236] If

---

[233] See id. at 1212.
[234] Fair Credit Reporting Act, 15 U.S.C. §§1681-1681u (2000).
[235] See Mitchell, supra note 13, at 83-119.
[236] See Rachlinski, supra note 4, at 620-21; supra notes 4-5 and accompanying text.

so, then one will have altered some legal outcomes that were not in any need of reform at all (the outcomes that would have resulted from decisions by actors not suffering from hindsight bias). The contrast with strategies for debiasing through law is clear. If the government provides accurate information about alcohol usage on a given campus, those whose prior impression coincided with the actual facts are quite unlikely to alter their drinking behavior.

C. New Distortions

While debiasing through substantive law can be analogized to more conventional government correctives to information failures, it is clear that some strategies for debiasing through law pose greater normative challenges and may create some significant dangers. As we have already suggested, one set of problems arises when the government intervention has significant effects on those whose behavior did not exhibit bounded rationality prior to the intervention. Consider, for example, measures that themselves harness other departures from bounded rationality, as with the strategic employment of the availability heuristic in response to optimism bias on the part of consumers, discussed in Part II.A above. In such cases, the legal intervention may produce affirmative distortions in the behavior of individuals who did not exhibit any bounded rationality in the first place. For those who previously had an accurate understanding of the situation, the debiasing effort could produce a kind of unrealistic pessimism. In such cases, it is no longer possible to say that, even if the legal intervention does not provide much help, it is unlikely to cause much harm. The same problem arises with respect to debiasing through corporate law in the form of the Sarbanes-Oxley Act, discussed in Part II.F above; the presence of outsiders on the board typically imposes a range of costs and thus has real potential to affect insiders who did not suffer from optimism bias as well as those who did.

The problem of adverse effects on actors who did not previously err will be reduced to the extent that those who are not the targets of a debiasing through law strategy are also less likely to be influenced by heuristics and biases harnessed by that strategy. Suppose, for example, that consumers who are relatively immune to optimism bias will not be led by a government strategy harnessing availability – and thus calling for concrete accounts of individual harm – to overestimate the probability of harm. If so, then that strategy will not run the risk of affirmatively engendering bias in individuals who previously did not exhibit departures from unbounded rationality.

Empirical investigation would be crucial in determining which sorts of interventions might and might not lead previously unbiased individuals astray.

Does the possibility of effects on those not otherwise prone to error mean that debiasing through law should be avoided in such circumstances? That conclusion would be too extreme. The question is the aggregate effect. Our point in the preceding subsection was that often there is no likelihood of significant effects on those not prone to error, and intervention in those cases seems most likely to be unobjectionable. But even outside of those situations, if debiasing through law improves accuracy for the large majority of people, it may be desirable even if it decreases accuracy for a few. Here, as in other contexts, the only option is to weigh the effects of the different possible strategies. Of course efforts to debias people through law should be undertaken, if possible, in a way that does not produce confusion or misperception.

D. Overshooting and Autonomy

Debiasing through law strategies that themselves harness other departures from bounded rationality – as with the employment of the availability heuristic in response to optimism bias – raise two independent concerns. One is the risk of overshooting. If truthful narratives or reframing strategies are used, people who previously exhibited optimism bias might be led to exaggerate the risks of consumer products. The behaviorally informed effort at debiasing through law would then be producing biases and errors of its own. Experimentation would again be required to calibrate correctly the degree to which availability or framing effects would need to be brought to bear – just as, in a conventional "debiasing law" approach, experimentation is necessary to determine the appropriate level or scope of the legal response. As already noted, however, debiasing through law strategies – unlike their "debiasing law" counterparts – at least preserve the option that the legal reforms will have only limited effects on those who did not previously err.

Overshooting may also arise through the operation of group polarization, discussed in Part II.F above in the context of corporate boards. Standing alone, the introduction into a group setting of one or more unbiased individuals would presumably not present a real risk of overshooting. But a critical mass of unbiased members alongside the preexisting optimists could conceivably polarize in an unduly pessimistic direction. Here as well, there is

an important role for experimentation to determine the appropriate degree of correction.

A second and more fundamental concern with debiasing through law involves the interest in autonomy. In some cases of such debiasing, government seems to be correcting bounded rationality by exploiting it, in a way that might give rise to fears of manipulation. In our examples in Part II, this occurs most obviously with respect to harnessing availability and framing in response to optimism bias on the part of consumers. Is this a legitimate form of government action? If heuristics and biases are pervasive, then an informed government is likely to have little trouble in manipulating people in its preferred directions. The problem here is that government should respect its citizens, as contemplated, for instance, by the publicity condition in John Rawls's *A Theory of Justice*[237]: Government should not engage in acts that it could not defend in public to those who are subject to those acts. If a public defense could not be made, the acts are an insult to the autonomy of citizens. The publicity condition raises some particular questions about any governmental effort to enlist bounded rationality in its preferred directions; and the concern about manipulation is a broader one still.

If manipulation itself is the focus, the initial response to the autonomy concern is that, from a behaviorally informed perspective, the worry about government manipulation arises even with the widely accepted approach under which the government corrects simple information failures among citizens. As is implicit in much of the discussion above, there is usually no neutral way to present information. Whenever the government is presenting even accurate information, it is making choices about its presentation, choices that will affect how citizens perceive the reality around them. (Public employees who have been subject to retirement options will easily recognize the point.[238]) Thus, it is far too simple, and behaviorally naive, to draw a sharp line between acceptable "provision of information" and unacceptable "mind control." Unless the concern with government manipulation is strong enough to suggest that the government should never provide information to its citizens (an implausible suggestion), there must be some willingness to tolerate the prospect of government influence over citizens' perceptions of

---

[237] John Rawls, A Theory of Justice 133 (1971).

[238] See Shlomo Benartzi & Richard H. Thaler, Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments, 45 Mgmt. Sci. 364, 375-77 (1999).

reality and the attendant risk of government manipulation. Thus, for example, if smokers discount the risks that accompany smoking, in part because of optimism bias, it is anything but obvious that government violates their autonomy by giving a more accurate sense of those risks, even if the best way of give that accurate sense is through vivid portrayals of suffering. And it is far from clear in such a case that the government could not publicly defend its strategy to citizens; recall in this connection the American Legacy Foundation letters campaign described above.[239]

This is not to say of course that all conceivable forms of debiasing through law (substantive or otherwise) would be unobjectionable on grounds of government manipulation. Some forms of such debiasing might resemble systems of propaganda in clear violation of the publicity condition. In counteracting underage alcohol use, for instance, some of those who have tried to debias students have gone far beyond anything plausibly described as a presentation of the facts, or even the use of a discrete heuristic, in pursuit of their efforts. Some attempts have involved large-scale public advertising campaigns, complete with campus posters, use of public computer terminals, and classroom intervention, and have employed self-consciously one-sided use of information.[240] As the leading advocate of this sort of approach has revealingly and somewhat chillingly put the point: "If one measure of an actual norm is not as positive as we might like, we should consider . . . what other measures might also be available that give a different picture."[241] One step in an illustrative campaign (and this step by itself might be fairly innocuous) involved campus posters; a first set, called "Reality Check," began with a description of widely believed myths and then offered corrections, while another set, called "Healthy Choices Are on the Rise," described recent increases in the number of students who did not miss class or engage in risky sexual practices as a result of drinking.[242] As noted, campus computers were also enlisted, as library and administrative computers displayed relevant messages whenever they remained idle for ten minutes. [243] The curriculum was affected as well, with a team-taught course on alcohol use and abuse, and with some general discussion of the social norms campaign in the classroom.[244]

---

[239] See supra note 102.
[240] See Perkins & Craig, supra note 171, at 40-51.
[241] Perkins, supra note 161, at 10.
[242] Perkins & Craig, supra note 171, at 40-43.
[243] Id. at 43-48.
[244] Id. at 48-51.

Under approaches of this sort, there is a real risk that the one-sidedness and aggressiveness of the effort will be exposed and therefore will reduce trust. And if trust is reduced, strategies for debiasing through law are much less likely to succeed. These instrumental concerns are aggravated by moral ones: At least when minors are not involved, the law should treat citizens with respect, and extreme marketing strategies of this sort (going well beyond what we have suggested in the consumer safety context) violate that principle. Compare imaginable efforts to control sex discrimination through detailed requirements for public portraits of women or through controlling people's use of sexually explicit pictures in their homes. At the very least, such efforts might be counterproductive, although it is important to emphasize that the area of discrimination is distinctive, and preference-shaping may be more acceptable in this context than in others. But even here it would be possible to fear manipulation.

We are not able to reach any general conclusion about the normative issues associated with debiasing through law in situations in which overshooting and threats to autonomy are of concern. In fact no general conclusion is likely to make sense; the strength of instrumental objections to strategies for debasing through law on grounds of overshooting depends on the setting, and the same is true for moral objections to public manipulation. Both sets of objections seem weakest when government is responding, as in the consumer context, to an identifiable bias and is using methods that do not distort the facts. Our hope is that the contextual discussion of particular areas in Part II and our development of the distinctive normative issues raised by these areas provide a basis for judgments about the strategies for debiasing through law we have examined.[245]

Nothing said thus far is meant to deny the fact that legal policymakers and administrators, including those who seek to engage in debiasing through

[245] Our normative analysis has omitted one of the examples of debiasing through law outlined in Part II: debiasing through the choice of liability rules over property rules. As noted in Part II.E, the context of the endowment effect – at issue in the discussion of property rules versus liability rules – raises normative issues distinct from those that arise in the context of judgment biases. For that reason our analysis in Part II.E. simply posited a normative desire to bring willingness to accept down to the level of willingness to pay. With that assumed normative goal, there are no apparent issues with our suggested means.

law, will often suffer from both inadequate information and bounded rationality. No less than ordinary people, bureaucrats themselves use heuristics and are subject to predictable biases; in addition, they are susceptible to the influence of powerful private groups with stakes in the outcome. The combination of cognitive biases and interest-group power can lead government in extremely unfortunate directions. In this light we do not make the naive and implausible suggestion that in the real world, debiasing strategies will always be well-motivated and well-designed. Our claim is only that if people make mistakes as a result of bounded rationality, and therefore reduce their own welfare, debiasing through law is often the most promising response. It would be foolish to eliminate that response from government's repertoire, especially because the most prominent alternatives are far more intrusive and at least equally subject to abuse.

Conclusion

The central goal of this Article has been to draw attention to the possibility of debiasing through substantive law. The social science literature has devoted a great deal of effort to the general study of debiasing of boundedly rational actors, but largely through the provision of information by experts, and with no effort to see how law and legal institutions might accomplish this goal. Those interested in bounded rationality and law have argued mostly that legal institutions should be insulated from the effects of boundedly rational behavior, and in some cases that debiasing should be pursued through changes in procedural rules. In our view, debiasing, especially debiasing through substantive law, is a distinctive and sometimes far preferable alternative to the strategy of attempting to insulate legal outcomes from the effects of bounded rationality. In many contexts, debiasing through substantive law promises to be both more successful and less invasive than the more familiar alternatives. Sometimes consumers are too optimistic; the availability heuristic and reframing might be enlisted as correctives. Employment decision makers often show unconscious bias, and steps could be taken – indeed, some already have been – to reduce this bias. Debiasing through government information campaigns, changing default rules, and the structure of property and corporate law are also illustrations.

From the normative point of view, many strategies for debiasing through law belong in the same category as more familiar efforts to respond to information failures by providing additional facts. Indeed, debiasing through law may be seen as a distinctive kind of informational regulation. In

many cases, the major questions are standard: whether such efforts are effective and whether their benefits justify their costs. But some imaginable efforts at debiasing raise serious normative questions and (if they are directed at speech) constitutional problems as well. We have emphasized that government should not regulate speech because it thinks that people are likely to be influenced by it. In the cases that we have discussed, however, government's target is unlawful conduct, and a form of "debiasing" generally accompanies efforts to forbid such conduct, even in the most uncontroversial domains of civil and criminal law. In that sense, law is pervasively in the business of debiasing.

Nothing in our analysis is inconsistent with the claim that in some contexts unfettered markets are the best response to bounded rationality. Such markets might reduce the effects of bounded rationality by raising the stakes, as noted above[246]; it is also possible that the costs of boundedly rational behavior are, in some contexts, lower than the costs of any effort to counteract it. We also do not disagree with the now-familiar suggestion that in the face of bounded rationality, aggressive regulation – some form of "debiasing law" – might sometimes be justified.[247] Instead our goal in this Article has been to chart the possibility of a middle course, one that asks legal institutions not to ignore people, but instead to reduce their errors. In some contexts, debiasing of boundedly rational actors is likely to be effective, cost-justified, and minimally intrusive. We believe that numerous areas of the law reveal an appreciation of these points and hence an implicit behavioral rationality, using legal strategies as a mechanism for debiasing of boundedly rational individuals. Our principal goal has been to understand those strategies in these terms, and to explore the possibility of building on them to do far more.

[246] But for an entertaining demonstration of the persistence of bounded rationality amidst high stakes, see Michael Lewis, Moneyball (2003).
[247] See Jolls, Sunstein, & Thaler, supra note 1, for many examples.

Appendix: Debiasing Strategies

This appendix is designed to summarize empirical research on debiasing and potential legal applications. In some cases, there is little evidence on whether debiasing is successful or unsuccessful; and in other cases, the evidence is disputed. Our emphasis here has been on debiasing through law, not debiasing law; we include examples of the latter (1 and 2 below) for purposes of comparison.

	Evidence of unsuccessful debiasing?	Evidence of successful debiasing?	Legal application
1. Hindsight bias	Almost all approaches Source: Baruch Fischhoff, Debiasing, in Judgment under Uncertainty 422, 424 (Daniel Kahneman et al. eds., 1982)	Tell people not to be "Monday Morning Quarterbacks" in closing arguments. Source: Merrie Jo Stallard & Debra L. Worthington, Reducing the Hindsight Bias Utilizing Attorney Closing Arguments, 22 J. L. & Hum. Behav. 671, 680-81 (1998).	Business judgment rule
2. Self-serving bias	Alerting people to existence of the bias Source: Linda Babcock, George Loewenstein, & Samuel Issacharoff, Creating Convergence: Debiasing Biased Litigants, 22 L. & Soc. Inquiry 913 (1997)	Asking people to make other side's argument Source: Linda Babcock, George Loewenstein, & Samuel Issacharoff, Creating Convergence: Debiasing Biased Litigants, 22 L. & Soc. Inquiry 913 (1997)	Damage caps
3. Unrealistic optimism	Almost all approaches that directly target the unrealistic optimism Source: Neil D. Weinstein and William Klein, Resistance of Personal Risk Perceptions to Debiasing Intervention, in Heuristics and Biases: The Psychology of Intuitive	Give people a vivid sense of the relevant harms Source: Frank A. Sloan, Donald H. Taylor, & V. Kerry Smith, The Smoking Puzzle: Information, Risk Perception, and Choice 122-23, 127, 161 (2003).	1. Consumer protection: Use availability heuristic and framing to counteract excessive optimism 2. Corporate law: outside directors on

	Judgment 313, (Thomas Gilovich et al. eds. 2003)		corporate boards
4. Unconscious racial bias		Show picture of admired people from relevant groups before the test Source: Nilanjana Dasgupta & Anthony G. Greenwald, On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals, 81 J. Personality & Soc. Psych. 800, 803-04 (2001).	Title VII reforms designed to reduce unconscious bias
5. Availability heuristic and associated biases		Give people accurate information about probabilities Source: H. Wesley Perkins, The Social Norms Approach to Preventing School and College Age Substance Abuse 7-8 (2003).	Social norms approaches to binge drinking on college campus, emphasizing actual number of people involved in binge drinking
6. Legal fairness heuristic (believing that law fits judgments about what is fair)	Experience in workforce or in unions Source: Pauline Kim, Norms, Learning, and Law, 1999 U. Ill. L. Rev. 447, 451.		Shift default rule so that workers have an accurate perception of legal rule
7. Endowment effect	Almost all approaches Source: Daniel Kahneman, Jack Knetsch & Richard Thaler, Experimental Tests of the Endowment Effect and	Protect entitlements via liability rules rather than property rules Source: Jeffrey J. Rachlinski & Forest Jourden, Remedies and the Psychology of	Use liability rules rather than property rules

	the Coase Theorem, 98 J. Pol. Econ. 1325 (1990).	Ownership, 51 Vand. L. Rev. 1541 (1998).	